

# SEARCH REQUEST FORM

Requestor's  
Name: \_\_\_\_\_

Serial  
Number: \_\_\_\_\_

Date: \_\_\_\_\_

Phone: \_\_\_\_\_

Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 4/11

Searcher: D. Schuster 308-4292

Terminal time: 8

Elapsed time: 7

CPU time: \_\_\_\_\_

Total time: \_\_\_\_\_

Number of Searches: \_\_\_\_\_

Number of Databases: 4

### Search Site

\_\_\_\_\_ STIC

☒ CM-1 6403

\_\_\_\_\_ Pre-S

### Type of Search

4 N.A. Sequence

2 A.A. Sequence

\_\_\_\_\_ Structure

\_\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_\_ IG

\_\_\_\_\_ STN

\_\_\_\_\_ Dialog

\_\_\_\_\_ APS

\_\_\_\_\_ Geninfo

\_\_\_\_\_ SDC

\_\_\_\_\_ DARC/Questel

☒ Other Computer

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

















DB 198 GGTCAATGATGATG 185

## RESULT 15

LOCUS H45410 430 bp mRNA EST 31-JUL-1995

DEFINITION y066605.s1 Soares breast 3dhbmat Homo sapiens cDNA clone

IMAGE:182864.3', mRNA sequence.

ACCESSION H45410.1 GI:921462

## KEYWORDS

EST.

## ORGANISM

human.

## REFERENCE

1 (bases 1 to 430)

Authors: Soares, M., Tan, F., Trevisan, E., Waterson, R., Williams, A., Mohdamm, P., and Wilson, R.

Title: Unpublished EST Project

Journal: Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 645

HiQ size: 319

Source: IMAGE Consortium

This clone is available royalty-free through LHM: contact the

IMAGE Consortium (info@image.lhm.gov) for further information.

Insert Length: 645 Std Error: 0.00

Seq primer: Promega -21m3

High quality sequence stop: 319.

Location/Qualifiers

1..430

1..430

1..430

1..430

1..430

1..430

1..430

1..430

1..430

1..430

1..430

1..430

1..430

## RESULT 16

LOCUS BC213986 431 bp mRNA EST 21-APR-2001

DEFINITION R3733612 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

ACCESSION BC213986.1 GI:13735673

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

human.

## REFERENCE

1 (bases 1 to 431)

Authors: Soares, M., Tan, F., Trevisan, E., Waterson, R., Williams, A., Mohdamm, P., and Wilson, R.

Title: Unpublished EST Project

Journal: Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 645

HiQ size: 319

Source: IMAGE Consortium

This clone is available royalty-free through LHM: contact the

IMAGE Consortium (info@image.lhm.gov) for further information.

Insert Length: 645 Std Error: 0.00

Seq primer: Promega -21m3

High quality sequence stop: 431.

Location/Qualifiers

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

## RESULT 17

LOCUS AM26301 312 bp mRNA EST 09-JUL-2000

DEFINITION 18672 Sus scrofa 2p1g Sus scrofa cDNA 5', mRNA sequence.

ACCESSION AM26301.1 GI:6762222

## KEYWORDS

EST.

## SOURCE

pig.

## ORGANISM

Sus scrofa

## REFERENCE

1 (bases 1 to 312)

Authors: Soares, M., Tan, F., Trevisan, E., Waterson, R., Williams, A., Mohdamm, P., and Wilson, R.

Title: Unpublished EST Project

Journal: Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 645

HiQ size: 319

Source: IMAGE Consortium

This clone is available royalty-free through LHM: contact the

IMAGE Consortium (info@image.lhm.gov) for further information.

Insert Length: 645 Std Error: 0.00

Seq primer: Promega -21m3

High quality sequence stop: 431.

Location/Qualifiers

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431

1..431









Query Match 1.28: Score 59: DB 11: Length 279:  
 Best Local Similarity 100.0%; Pred. No. 67e-07;  
 Matches 59: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 4475 gaagtaactgcattatccgacacacacctctttagtcgcgcagaggtacacagatga 4533  
 |||||||  
 Db 28 GCAATTATTCGCAATTTCGCAACACCTTCTTAGTCTGCACAGGCTCACATGATGATGA 86

RESULT 25  
 AA502357/c 421 bp mRNA EST 19-AUG-1997  
 LOCUS n62606.n1.NCI\_CGAP.C03 Homo sapiens cDNA clone IMAGE:898474.3',  
 DEFINITION n62606.n1.NCI\_CGAP.C03 Homo sapiens cDNA clone IMAGE:898474.3',  
 ACCESSION AA502357  
 VERSION AA502357.1 GI:2237324  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 (bases 1 to 421)  
 NC\_000001: www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: CGAPdb-femail.nih.gov  
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D., Library Preparation: M. Benito Soares, Ph.D.  
 DNA Library: Arthur J. Schram, Ph.D., Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/biopr/image/image.html  
 Insert Length: 529 Std Error: 0.00  
 Seq primer: -40m13 fwd. RT from Amersham  
 High quality sequence stop: 395.  
 Location/Qualifiers  
 1..421  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:898474"  
 /clone\_lib="NCI\_CGAP.C03"  
 /sex="pooled"  
 /lab\_host="EMBL09"  
 /map\_pos="1000000"  
 /note="Vector: pRT73-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI. A 1.2 kb fragment  
 was prepared from 12 pooled bulk tumor samples and subcloned  
 with a Not I - oligo(dT) primer. Double-distended cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 1.2 kb pRT73 vector. Library went through one round of  
 normal fractionation. Library went through one round of  
 normal fractionation. 83 g 110 t

BASE COUNT 132 a 96 c 110 t  
 ORIGIN

Query Match 1.18: Score 50: DB 10: Length 421:  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 50: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 4484 gcaattatgcacacacacctctttagtcgcgcagaggtacacagatga 4533  
 |||||||  
 Db 232 GCAATTATTCGCAACACCTTCTTAGTCTGCACAGGCTCACATGATGATGA 183

RESULT 26  
 AA594918 363 bp mRNA EST 25-SEP-2000  
 LOCUS AV692878  
 DEFINITION AV692878 GKC Homo sapiens cDNA clone GKCC2B10.5', mRNA sequence.

ACCESSION AV694918  
 VERSION AV694918.1 GI:10296781  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 (bases 1 to 363)  
 NC\_000001: www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (2000)  
 Contact: Zengqiang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzengqiang@sh.cn  
 This clone was sequenced at CHGC in Shanghai.  
 Location/Qualifiers  
 1..363  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="GKCC2B10"  
 /clone\_lib="GKC"  
 /sex="LYPE" Hepatocellular carcinoma"  
 /day="1997-04-18"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); site 1: EcoRI; site 2:  
 XhoI"

BASE COUNT 122 a 57 c 95 g 88 t 1 other  
 ORIGIN

Query Match 1.08: Score 47: DB 10: Length 363:  
 Best Local Similarity 100.0%; Pred. No. 0.0012;  
 Matches 47: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1058 agttcttactacagacagatggtagacacacgcgcacattcccaacctt 1104  
 |||||||  
 Db 145 AGCTTCTACTCTACAGATTCGCAACATTCGTCACATTCGCAACCTT 191

RESULT 27  
 AV692878 705 bp mRNA EST 25-SEP-2000  
 LOCUS AV692878 GKC Homo sapiens cDNA clone GKCCYD05.5', mRNA sequence.  
 ACCESSION AV692878  
 VERSION AV692878.1 GI:10294741  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 (bases 1 to 705)  
 NC\_000001: www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (2000)  
 Contact: Zengqiang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzengqiang@sh.cn  
 This clone was sequenced at CHGC in Shanghai.  
 Location/Qualifiers

SOURCE

1. 705  
/organism="Homo sapiens"  
/clone="H8606"  
/db\_xref="GKCYD05"  
/clone.lib="GKC"  
/tissue.type="hepatocellular carcinoma"  
/dev.stage="Adult"  
/lab.host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 247 a 111 c 165 g 181 t 1 others  
ORIGIN

Query Match 1.0%; Score 47; DB 10; Length 705;  
Best Local Similarity 100.0%; Pred. No. 0.00074;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1058 agttctactatgcagaattggagcaatcctggacattcttaacctt 1104  
|||||  
DB 137 agttctactatgcagaattggagcaatcctggacattcttaacctt 183

RESULT 28  
AV690671 754 bp mRNA EST 25-SEP-2000  
LOCUS AV690671 GKC Homo sapiens cDNA clone GKCYH03 5', mRNA sequence.  
DEFINITION AV690671  
ACCESSION AV690671.1 GI:10292534  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 754)  
Wu,T., Qian,B., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,  
Xu,S., Gu,W., Tu,Y., Jia,J., Qu,J., Song,H., Cheng,Z., Zeng,L.,  
Xia,S., Gu,W., Tu,Y., Jia,J., Qu,J., Ren,S., Zhong,X., Yu,G., Yang,  
H., et al. (1999) Human genome project. Nature 397:121-122.  
KEYWORDS  
UNPUBLISHED (2000)  
CONTACT: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex. 45)  
Fax: 86-21-50801923  
E-mail: zhengz@cnhg.ac.cn  
This clone is available at CHGC in Shanghai.  
LOCATION/Qualifiers  
1. 754  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone.lib="GKC"  
/clone.lib="GKC"  
/tissue.type="hepatocellular carcinoma"  
/dev.stage="Adult"  
/lab.host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI"

FEATURES

source  
BASE COUNT 243 a 125 c 175 g 211 t  
ORIGIN

Query Match 1.0%; Score 47; DB 10; Length 754;  
Best Local Similarity 100.0%; Pred. No. 0.0007;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1058 agttctactatgcagaattggagcaatcctggacattcttaacctt 1104  
|||||  
DB 139 agttctactatgcagaattggagcaatcctggacattcttaacctt 185

RESULT 29

AC0893428/c 828 bp DNA GSS 10-NOV-1999  
LOCUS AC0893428 H8\_3100\_B2\_H11\_77C C19 Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone plate-3100 Col-22 Row-P, DNA sequence.  
ACCESSION AC0893428.1 GI:6349618  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 828)  
Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,W.D. and  
Hood,L. (2000) A sequence approach to mapping and  
sequencing the human genome.  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
CONTACT: Mahatras GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3618  
E-mail: mahatras@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 3100 Row: P Column: 22  
Seq primer: 47  
Class: BAC ends  
High quality sequence stop: 828.  
LOCATION/Qualifiers  
1. 828  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone.lib="C19 Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: Sperm; Vector: pBluescript II; BAC Clones In E-Clon DH10B"

FEATURES

source  
BASE COUNT 272 a 147 c 148 g 261 t  
ORIGIN

Query Match 1.0%; Score 47; DB 13; Length 828;  
Best Local Similarity 100.0%; Pred. No. 0.00065;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1058 agttctactatgcagaattggagcaatcctggacattcttaacctt 1104  
|||||  
DB 617 agttctactatgcagaattggagcaatcctggacattcttaacctt 571

RESULT 30  
BG253578 883 bp mRNA EST 13-FEB-2001  
LOCUS BG253578  
DEFINITION 603364324F1 NIH\_MGC\_50 Homo sapiens cDNA clone IMAGE:447692 5',  
mRNA seq. sequence.  
ACCESSION BG253578.1 GI:12763394  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 883)  
NIH-MGC (http://www.ncbi.nlm.nih.gov/  
NIH-MGC) National Institutes of Health, Mammalian Gene Collection (MGC)  
UNPUBLISHED (1999)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapds@mail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMLL at:  
<http://image.llnl.gov>  
 Plate: LLM10234 row: p column: 05  
 High quality sequence stop: 690.

Location/Qualifiers  
 1..883

## FEATURES

## source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="4472692"  
 /clone\_image="444967"  
 /tissue\_type="adipocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Liver; Vector: pCMV-SPORT6; Site:1: NOTI;  
 Site:2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: This is a NIH-MGC Library."  
 BASE COUNT 305 a 159 c 207 g 212 t  
 ORIGIN

Query Match 1.0%; Score 47; DB 11; Length 883;  
 Best Local Similarity 100.0%; Pred. No. 0.00062;  
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1058 agttcttctacagagattggtgcacatctgcacatctcaccctt 1104  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 26 AGTCTTACTGTCACACATTCGTGCACATCTGCACATTCACACTT 72

RESULT 31  
 LOCUS 3584  
 DEFINITION BGI21584 929 bp mRNA EST 30-JAN-2001  
 accession 602351531f NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:444967 5',  
 mRNA sequence.  
 BGI21584  
 VERSION BGI21584.1 GI:12615093  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 929)  
 AUTHORS NIH-MGC <http://img.ncl.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgsapb@remail.nih.gov  
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LMLL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMLL at:  
<http://image.llnl.gov>  
 Plate: LLM10234 row: m column: 08  
 High quality sequence stop: 722.

## FEATURES

## source

Location/Qualifiers  
 1..329  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="444967"  
 /clone\_image="444967"  
 /tissue\_type="adipocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site:1: NOTI;  
 Site:2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.  
 Note: This is a NIH-MGC Library."  
 BASE COUNT 330 a 139 c 247 g 213 t  
 ORIGIN

Query Match 1.0%; Score 47; DB 11; Length 929;  
 Best Local Similarity 100.0%; Pred. No. 0.00062;  
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1058 agttcttctacagagattggtgcacatctgcacatctcaccctt 1104  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 239 AGTCTTACTGTCACACATTCGTGCACATCTGCACATTCACACTT 285

RESULT 32  
 LOCUS 244 bp mRNA EST 21-APR-2001  
 DEFINITION BGI92713 244 bp mRNA EST 21-APR-2001  
 accession BGI92713.1 GI:13724400  
 VERSION BGI92713.1 GI:13724400  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 244)  
 AUTHORS Hall, J. D., Shaf, B., Rundlett, S., Jackson, P. D., Perry, R.,  
 Calkins, D., Shaf, B., Shaf, B., Shaf, B., Shaf, B., Shaf, B.,  
 Lerner, L., Krause, D., McMillan, K., Clapham, R., Whitlington, J.,  
 Veloso, N., Hess, J., Coltrien, K., Lo, K., Offenbacher, J., Smith, D.,  
 and Ducar, M.  
 TITLE Creation of Genome-wide Protein Expression Libraries using Random  
 Activation of Gene Expression  
 JOURNAL Nature Biotechnology, 19 (5), 440 (2001) In press  
 COMMENT Contact: Scott J. Cain  
 Address: 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scalc@case.edu  
 High quality sequence stop: 194.

Location/Qualifiers  
 1..244

## FEATURES

## source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="444967"  
 /clone\_image="444967"  
 /tissue\_type="adipocarcinoma, cell line"  
 /note="See 'Creation of Genome-wide Protein Expression  
 Libraries using Random Activation of Gene Expression',  
 Nature Biotechnology, In press. Note that even though the  
 cell type indicated is H1080, since a random activation  
 method was used, these sequence tags are not necessarily  
 expressed in H1080 under normal circumstances."  
 BASE COUNT 44 a 82 c 57 g 59 t  
 ORIGIN

Query Match 0.9%; Score 43; DB 11; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 0.022;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4488 ttatcgacacacatctttagcttcacacaggtcacatgca 4530  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 94 TTATCGACACACATCTTATAGCTTCACACAGGTCAATGCA 136

RESULT 33  
 LOCUS 171 bp mRNA EST 02-DEC-1997  
 DEFINITION A680196.1 Strausberg Lung (4937210) Homo sapiens cDNA clone  
 IMAGE:869071 3', mRNA sequence.  
 A680196

VERSION	AA680196.1	GI:2656164
KEYWORDS	EST.	
SOURCE	Human.	
ORGANISM	Homo sapiens	
REFERENCE	National Human Genome Research Institute; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	I (bases 1 to 171) Hillier,L., Allen,M., Bowles,L., Dubnue,T.G., Gelisell,G., Jost,S., Kliman,D., Kuenz,M., Powell,L., Lacey,M., Le,N., Lennon,G., Maiz,M., Martin J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theisling,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R. Washington University School of Medicine Box 8501, St. Louis, MO 63108	
TITLE	WashU-NCI Human EST Project	
JOURNAL	Unpublished	
COMMENT	Submitted to GenBank Contact: Wilson RK 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LLN <sup>*</sup> ; contact the LLN at llncbi@wustl.wustl.edu or visit the website at http://www.llnwd.com/submit/submit.asp for further information. IMBOL Consortium, NCI and NIDDK from Amsterdam High quality sequence stop: 159.	
FEATURES	Location/Qualifiers	
source	1..171 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:66907" /feature="11bp" Strandage Lung (#937210)" /sex="male" /dev_stage="72 years" /lab_host="SOLR cells (kanamycin resistant)." /note="Organ: lung; Vector: pBluescript SK+; Site: 1; ECORI site, Site:2; XhoI; Cloned unidirectionally." Primer: Oligo dr. normal lung. Average insert size: 2,000 bp Vector: 3' adapt. sequence: 5'-GGGAGCGACGCGCAG-3' adapt.3' sequence: 5'-CTCCAGCTTTTTTTTTTTTTTT 3'-R 29 a 5g 57 t.	
BASE COUNT	a 29 g 5g 57 t.	
ORIGIN		
Query Match	0.9%; Score 41; DB 10; Length 171;	
Best Local Similarity	100.0%; Pctd. No. 0.1;	
Matches	41; Conservative 0; Mismatched 4500	
Oy	4460 gccagccgcacgaagatcgaatttcgcatctaccgacaca 4500 	
Db	41 GACGACGACCACCAAGAGTAATTCGCAATTATTCGGACACA 1	
RESULT 34		
BFB80177	354 bp mRNA EST	17-JAN-2001
DEFINITION	CX2-ET0127-291100-524-a06 ET0127 Homo sapiens CDNA, mRNA sequence.	
ACCESSION	BFB80177	
VERSION	BFB80177.1	GI:1270303
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Muller-Hartmann; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 354)	
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Negel,M.A., da Silva,M.J., Zagzo,M., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.P.S., Matsukuma,A., Balda,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Buchner,P., Jongeneel,C.V., d'Oate M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.U. and Shounmura	
TITLE	Shoun sequencing of the human transcritome with ORF expressed sequence tags	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	
MEDLINE	20202663	
COMMENT	Contact: Simpson A.J.G.	

```

FEATURES
source
    Laboratory of Cancer Genetics  

    Ludwig Institute for Cancer Research  

    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  

    Brazil  

    Tel.: +55-11-27046922  

    Fax: +55-11-2707400  

    Email: genomac@ludwig.org.br  

    This sequence was derived from the FAPESP/LICR Human Cancer Genome  

    Project. This entry can be seen in the following URL:  

http://www.ludwig.org.br/scripts/gethtml2.pl?file=CM2&C=CM2-ET0127-291100-524-a06&L=2000-11-29&A=1  

    Seq primer: puc 18 forward  

    High quality sequence start: 21  

    High quality sequence stop: 350.  

    Location/Qualifiers
        1..1500
            /organism="Homo sapiens"  

            /db_xref="taxon:9606"  

            /clone_lib="ET0127"  

            /dev_stage="Adult"
        note="Organ: Lung; tumor; Vector: puc18; Site_1: SmaI;  

        Site_2: SmaI; A mini-library was made by cloning products  

        derived from GROMES PCR (U.S. Letters Patent application  

        No. 136,716 - Ludwig Institute for Cancer Research)  

        proteins mRNA and cDNA amplification were performed under  

        low stringency conditions."
        124 a          78 g          97 t
BASE COUNT      124 a          78 g          97 t
ORIGIN
Query Match           0.9%; Score 41; DB 11; Length 354;  

Best Local Similarity 100.0%; Pident. No. 0.05%;  

Matches 41; Conservative 0; Mismatched 0; Indels 0; Gaps 0;  

|||||  

|||  

Db 148 TATTGAGGTACCATTCATGGACCGAGCACTCCAG 1515  

TATTGAGGTACCATTCATGGACCGAGCACTCCAG 228

```







```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0487"
/dev_store="AduIt"
/notes="Organ: head;neck; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products
derived from OND57S PCR (U.S. Letters Patent application

```

BASE COUNT	172 a	70 c	102 g	91 t
ORIGIN				

Query Match	0.8%;	Score 37;	DB 10;	Length 435;
Best Local Similarity	100.0%;	Pred. No.	0.65;	





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: April 10, 2002, 14:58:41 ; Search time 3323.29 Seconds  
(without alignments)  
23480.244 Million cell updates/sec

Title: us-09-750-590-1

Reference score: 1 cagctgtgagcgcgagat.....atcaaaaaaaaaaaaaaa 4730

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: GenBank:\*

1: gb-act.\*

2: gb-act.\*

3: gb-act.\*

4: gb-act.\*

5: gb-act.\*

6: gb-act.\*

7: gb-act.\*

8: gb-act.\*

9: gb-act.\*

10: gb-act.\*

11: gb-act.\*

12: gb-act.\*

13: gb-act.\*

14: gb-act.\*

15: gb-act.\*

16: gb-act.\*

17: gb-act.\*

18: gb-act.\*

19: gb-act.\*

20: gb-act.\*

21: gb-act.\*

22: gb-act.\*

23: gb-act.\*

24: gb-act.\*

25: gb-act.\*

26: gb-act.\*

27: gb-act.\*

28: gb-act.\*

29: gb-act.\*

30: gb-act.\*

31: gb-act.\*

32: gb-act.\*

33: gb-act.\*

34: gb-act.\*

35: gb-act.\*

36: gb-act.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	1.6	4428	AF322916	AF322916 Homo sapi
2	74	1.6	4428	AF322916	AF322916 Homo sapi
3	74	1.6	4428	AF322916	AF322916 Homo sapi
4	61	1.3	15876	AC010076	AC010076 Homo sapi
5	45	1.0	4643	CECVS	CECVS Homo sapi
6	37	0.8	371	AX071367	AX071367 Sequence
7	37	0.8	401	AX071358	AX071358 Sequence
8	37	0.8	401	AX071358	AX071358 Sequence
9	37	0.8	401	AX071358	AX071358 Sequence
10	37	0.8	401	AX071358	AX071358 Sequence
11	35	0.7	441	AX193084	AX193084 Sequence
12	35	0.7	441	AX193084	AX193084 Sequence
13	32	0.7	421	AX193084	AX193084 Sequence
14	32	0.7	421	AX193084	AX193084 Sequence
15	32	0.7	421	AX193084	AX193084 Sequence
16	32	0.7	421	AX193084	AX193084 Sequence
17	29	0.6	83775	AC022466	AC022466 Homo sapi
18	28	0.6	1117	AB002533	AB002533 Homo sapi
19	24	0.5	2703	BOVPS	BOVPS Homo sapi
20	24	0.5	40534	AC005305	AC005305 Homo sapi
21	24	0.5	40534	AC005305	AC005305 Homo sapi
22	24	0.5	40534	AC005305	AC005305 Homo sapi
23	24	0.5	40534	AC005305	AC005305 Homo sapi
24	24	0.5	40534	AC005305	AC005305 Homo sapi
25	24	0.5	40534	AC005305	AC005305 Homo sapi
26	24	0.5	40534	AC005305	AC005305 Homo sapi
27	24	0.5	40534	AC005305	AC005305 Homo sapi
28	24	0.5	40534	AC005305	AC005305 Homo sapi
29	24	0.5	40534	AC005305	AC005305 Homo sapi
30	24	0.5	40534	AC005305	AC005305 Homo sapi
31	24	0.5	40534	AC005305	AC005305 Homo sapi
32	24	0.5	40534	AC005305	AC005305 Homo sapi
33	24	0.5	40534	AC005305	AC005305 Homo sapi
34	24	0.5	40534	AC005305	AC005305 Homo sapi
35	24	0.5	40534	AC005305	AC005305 Homo sapi
36	24	0.5	40534	AC005305	AC005305 Homo sapi
37	24	0.5	40534	AC005305	AC005305 Homo sapi
38	24	0.5	40534	AC005305	AC005305 Homo sapi
39	24	0.5	40534	AC005305	AC005305 Homo sapi
40	24	0.5	40534	AC005305	AC005305 Homo sapi
41	24	0.5	40534	AC005305	AC005305 Homo sapi
42	24	0.5	40534	AC005305	AC005305 Homo sapi
43	24	0.5	40534	AC005305	AC005305 Homo sapi
44	24	0.5	40534	AC005305	AC005305 Homo sapi
45	24	0.5	40534	AC005305	AC005305 Homo sapi

## ALIGNMENTS

RESULT 1	AF322916	4428 bp	mRNA	PRI	15-FEB-2001
LOCUS	AF322916				
DEFINITION	Homo sapiens uveal autoantigen mRNA, complete cds.				
ACCESSION	AF322916				
VERSION	AF322916.1	GI:12240160			
KEYWORDS					
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 4428)				
AUTHORS	Nakamura, S., Ohno, S., Negi, R. and Nishimura, Y.				
TITLE	Identification of a novel autoantigen Uvch in patients with				
JOURNAL	Immunobiology. Res. Commun. 280 (4), 1169-1176 (2001)				
PIRMD	1162550				
REFERENCE	2 (bases 1 to 4428)				
AUTHORS	Yamada, K., Senju, S. and Nishimura, Y.				











repeat_c_region	/P/C_family="MER46B" complement(11709..11758)
repeat_c_region	/P/C_family="MIR" 11784..11861
repeat_c_region	/P/C_family="AluS" complement(11723)..12119
repeat_c_region	/P/C_family="AluB" 12320..12339
repeat_c_region	/P/C_family="TTGn" complement(12340..12457)
repeat_c_region	/P/C_family="AluB" complement(12563..13020)
repeat_c_region	/P/C_family="MTRC" complement(14030..14428)
repeat_c_region	complement(15874..15935)
repeat_c_region	/P/C_family="MIR" complement(15972..16163)
repeat_c_region	/P/C_family="MIR" complement(16259..16326)
repeat_c_region	/P/C_family="L2" complement(18147..18235)
repeat_c_region	/P/C_family="L1b3" complement(18309..18444)
repeat_c_region	/P/C_family="MER5A" 18510..18625
repeat_c_region	/P/C_family="MIR3" complement(18811..19056)
repeat_c_region	/P/C_family="MIR" complement(19114..19160)
repeat_c_region	/P/C_family="L1b2" 19782..19847
repeat_c_region	/P/C_family="MIR" complement(20054..20111)
repeat_c_region	/P/C_family="MER5B" 20112..20420
repeat_c_region	/P/C_family="AluSg" 20431..20714
repeat_c_region	/P/C_family="AluS" complement(20799..20968)
repeat_c_region	/P/C_family="MER5B" 20799..20968
repeat_c_region	/P/C_family="AluSg/x" 20972..21283
repeat_c_region	/P/C_family="AluSg" complement(21287..21327)
repeat_c_region	/P/C_family="MER5B" 21328..21785
repeat_c_region	/P/C_family="L2" 21822..21943
repeat_c_region	/P/C_family="MIR" 21944..22247
repeat_c_region	/P/C_family="AluSx" 22248..22311
repeat_c_region	/P/C_family="MIR" 22356..22576
repeat_c_region	/P/C_family="L1b3" 23107..com23383
repeat_c_region	/P/C_family="AluSg" 23685..23976
repeat_c_region	/P/C_family="AluY" 23976..24000

Query Match	1.00;	Score 47;	DB 9;	Length 172850;
Best Local Similarity	100.00;	Score 47;	DB 9;	Length 172850;

Matches	47	conservative	0	Mismatches	0	Indels	0
---------	----	--------------	---	------------	---	--------	---

**Oy**      1058    ggccctcaccatgcagaaattggtgcaacatctggacatcttaacct    1104  
         |||||  
**Db** 165207 AGTCTTACTATGCAAGAATTGGTGACAATCTGCACATTTCTAACCTT    165253  
         |||||

RESULT  
CFC3VS , 7

LOCUS		6463 bp	MRNA			20-OCT-2000
DEFINITION	Cnals Familialis mRNa					
ACCESSION	X99145					
VERSION	X99145.2	GI:10944717				
KEYWORDS	cds	protein.				
SOURCE						
ORGANISM	Cnals Familialis					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.						
REFERENCE						
AUTHORS	Wilkin,F., Savonet,V., Radulescu,A., Petermans,J., Dumont,J.B. and Maenhaut,C.					
TITLE	Identification and characterization of novel genes modulated in thymic lymphomas treated with methimazole and propylthiouracil					
JOURNAL	J Biol Chem.	271 (45), 28451-28457	(1996)			
MEDLINE	97067069					
REFERENCE						
AUTHORS	2 (bases 1 to 6463)					
TITLE	Savonet,V.					
JOURNAL	Direct Submission					
TITLE	Submitted (06-JUL-1996) V. Savonet, Institute of Interdisciplinary Research Ula, Building C (local c4-124), Campus Erasme, Route de Reynders 60 (1070 Brussels, BELGIUM					
REMARK	3 (bases 1 to 6463)					
REFERENCE	Savonet,V.					
AUTHORS	Direct Submission					
TITLE	Submitted (20-OCT-2000) V. Savonet, Institute of Interdisciplinary Research Ula, Building C (local c4-124), Campus Erasme, Route de Reynders 60 (1070 Brussels, BELGIUM					
JOURNAL	On Oct 21 2000 th					
COMMENT	location/Qualifiers					
FEATURES						
SOURCE	1..4663					

BASE COUNT	1645 a	872 c	1185 g	941 t
ORIGIN				

Query Match 1.0% Score 45; DB 4; Length 4643;  
 Best Local Similarity 100.0%; Pred. No. 1 2e-11;  
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 696 tctgtcacaacactctcagacacattcccccgaacatg 740  
 |||||||  
 Db 677 tctgtcacaacacactctcagacacattcccccgaacatg 721

RESULT 8  
 AX071367 371 bp DNA PAT 25-JAN-2001  
 LOCUS Sequence 1839 from Patent WO0102568.  
 DEFINITION AX071367  
 ACCESSION AX071367.1 GI:12581718  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Human; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 371)

REFERENCE  
 AUTHORS Williams L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klingner, J., Kassam, A., Reinhard, C., Randazzo, F., Kennedy, G.C., Pol, D., Lamson, G., Drmanac, R., Cikenjakov, R., Drmanac, S., Dickson, M., Labat, I., Leschkowitz, D., Kita, D., Garcia, V., and Strache-Grain, B.  
 TITLE Human genes and gene expression products  
 JOURNAL Patent: WO 0102568 A 1839 1 JAN 2001;  
 CHIRON CORPORATION, 10000 N. DE SOTO, INC. (US)  
 LOCATION/Qualifiers  
 1 - 371  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 107 a 79 c 98 g 87 t  
 ORIGIN

Query Match 0.8%; Score 37; DB 6; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 1 8e-07;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 885 gttagccacaatctgcacatgcgcatacaaga 921  
 |||||||  
 Db 187 GTAGCCACAAATATGTCACCTCGATAGTAGAGG 223

RESULT 9  
 AX071358 401 bp DNA PAT 25-JAN-2001  
 LOCUS Sequence 1830 from Patent WO0102568.  
 DEFINITION AX071358  
 ACCESSION AX071358.1 GI:12581709  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Human; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 401)

REFERENCE  
 AUTHORS Williams L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klingner, J., Kassam, A., Reinhard, C., Randazzo, F., Kennedy, G.C., Pol, D., Lamson, G., Drmanac, R., Cikenjakov, R., Drmanac, S., Dickson, M., Labat, I., Leschkowitz, D., Kita, D., Garcia, V., and Strache-Grain, B.  
 TITLE Human genes and gene expression products  
 JOURNAL Patent: WO 0102568 A 1839 1 JAN 2001;  
 CHIRON CORPORATION (US); HYSDO, INC. (US)  
 LOCATION/Qualifiers  
 1 - 401  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 114 a 85 c 107 g 93 t 2 others

## ORIGIN

Query Match 0.8%; Score 37; DB 6; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 1 8e-07;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 885 gttagccacaatctgcacatgcgcatacaaga 921  
 |||||||  
 Db 187 GTAGCCACAAATATGTCACCTCGATAGTAGAGG 223

RESULT 10  
 G36947/c 418 bp mRNA STS 01-JAN-1998  
 LOCUS SHC-56623 Human Homo sapiens STS cDNA, sequence tagged site.  
 DEFINITION G36947  
 ACCESSION G36947.1 GI:2734614  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Human; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 418)

REFERENCE  
 AUTHORS Myers, R.M.  
 TITLE Human STS (1997)  
 JOURNAL Unpublished (1997)  
 COMMENT

Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Fax: 4157259689  
 Email: myers@SHGC.Stanford.edu  
 Primer A: GAGCTCTGCGACGGCGACG  
 Primer B: GGACGGGCTTGAGGCTCTG  
 STS size: 167  
 PCR Profile:  
 Initial incubation: 95 degrees C for 10 minutes  
 Denaturation: 94 degrees C for 30 seconds  
 Annealing: 59 degrees C for 30 seconds  
 Extension: 72 degrees C for 30 seconds  
 PCR cycles: 30  
 Thermal Cycler: Perkin Elmer 9600

Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Amplifying Gold Polymerase: 0.07 units/uL  
 Total vol: 5 uL

Buffer: MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

Prepared with primer pairs derived from M69830 -- Unigene.  
 LOCATION/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

FEATURES  
 source  
 STS  
 primer\_bind 151..317  
 primer\_bind 83 a 92 c 65 g 178 t  
 BASE COUNT 83 a 92 c 65 g 178 t  
 ORIGIN



PCR Profile:  
 Initial Incubation: 94 degrees C for 90 seconds  
 Denaturation: 94 degrees C for 15 seconds  
 Annealing: 64 degrees C for 23 seconds  
 Polymerization: 72 degrees C for 30 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9600

Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Taq Polymerase: 0.5 units/uL  
 Total Vol: 10 uL

Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 20 mM  
 pH: 8.3

Prepared with primer pairs derived from T47625--Merck/Unist.  
 Location/Qualifiers  
 1..250  
 /organism="Homo sapiens"  
 STS  
 primer\_bind 22..169  
 primer\_bind 22..40  
 complement(149..169)  
 BASE COUNT 70 a 63 c 56 g 58 t 3 others  
 ORIGIN

Query Match  
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;  
 Matches 32: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4471 ccaaggagatgcaattatcgagaccac 4502  
 |||||||  
 DB 225 CCAGCAAGATGCAATTGCAACGCGACAC 194

RESULT 14  
 AX192628 421 bp DNA PAT 15-AUG-2001  
 LOCUS  
 DEFINITION Sequence 195 from Patent WO0149716.  
 ACCESSION AX192628  
 VERSION AX192628.1 GI:15210591  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 421)  
 Xu,J., Lodes,M.J., Secretist,H., Benson,D.R., Meagher,M.J.,  
 Stok,J.A., King,G.,Therapy and diagnosis of colon cancer and  
 methods for their use  
 Patents: WO 0149716-A,195 12-JUL-2001;  
 JOURNAL CORIXA CORPORATION (US)  
 Location/Qualifiers  
 1..421  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 109 a 101 c 125 g 86 t

Query Match  
 Best Local Similarity 100.0%; Score 32; DB 6; Length 421;  
 Matches 32: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 617 atccatgagatgataatcacacagaccac 648  
 |||||||

DB 346 ATACATGAGATGATTTACACGCGACAC 377

RESULT 15  
 AX193058 421 bp DNA PAT 15-AUG-2001  
 LOCUS  
 DEFINITION Sequence 625 from Patent WO0149716.  
 ACCESSION AX193058  
 VERSION AX193058.1 GI:15211014  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 421)  
 Xu,J., Lodes,M.J., Secretist,H., Benson,D.R., Meagher,M.J.,  
 Stok,J.A., King,G.,Therapy and diagnosis of colon cancer and  
 methods for their use  
 Patents: WO 0149716-A,625 12-JUL-2001;  
 JOURNAL CORIXA CORPORATION (US)  
 Location/Qualifiers  
 1..421  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 109 a 101 c 125 g 86 t

Query Match  
 Best Local Similarity 100.0%; Pred. No. 7.5e-05;  
 Matches 32: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 617 atccatgagatgataatcacacagaccac 648  
 |||||||

DB 346 ATACATGAGATGATTTACACGCGACAC 377

RESULT 16  
 AC022466 83775 bp DNA HTG 04-FEB-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 15 clone CTD-302KX4 map 15q23, LOW-PASS  
 ACCESSION AC022466  
 VERSION AC022466.1 GI:6899623  
 KEYWORDS HTG: HTGS\_PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 83775)  
 James,R., Dots,M., Dickhoff,R., Fleetwood,P., Harrison,G.,  
 Bloom,S., Dots,M., Dickhoff,R., Fleetwood,P., Harrison,G.,  
 James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.  
 and Hood,L.  
 Sequencing of human chromosome 15 D15S146-D15S117 region  
 Unpublished  
 2 (bases 1 to 83775)  
 James,R., Dots,M., Dickhoff,R., Fleetwood,P., Harrison,G.,  
 Bloom,S., Dots,M., Dickhoff,R., Fleetwood,P., Harrison,G.,  
 James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.  
 and Hood,L.  
 Direct Submission  
 of Washington, PO Box 353730, Seattle, WA 98195, USA  
 -----Genome Center  
 Center for Human Genome Sequencing Center  
 Contact: http://chroma.mbc.washington.edu/msg\_www  
 Web site: http://chroma.mbc.washington.edu  
 Contact: leetowen@washington.edu  
 -----Summary Statistics  
 Sequencing vector: PUC18; L08752  
 Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-Primer Big Dye: 10% of reads  
 Assembly Program: Phrap: version 0.990399

NOTE: This record contains 96 individual  
 sequencing reads that have not been assembled into  
 contigs. The reads are used to separate the reads  
 and the order in which they appear is completely  
 arbitrary. Low-pass sequence analysis is not  
 identifying clones that may be genetic and all  
 overlap relationships among clones to be deduced  
 However, it should not be assumed that this clone  
 will be sequenced to completion. In the event that  
 the record is updated, the accession number will  
 be preserved.

1 887: contig of 897 bp in length  
 888 gap of unknown length  
 1780: contig of 893 bp in length  
 1781 gap of unknown length  
 2645: contig of 865 bp in length  
 3525: contig of 880 bp in length  
 3526 gap of unknown length  
 4401: contig of 876 bp in length  
 4402 gap of unknown length  
 5290: contig of 889 bp in length  
 5291 gap of unknown length  
 6155: contig of 865 bp in length  
 6156 gap of unknown length  
 7151: contig of 996 bp in length  
 7152 gap of unknown length  
 8036: contig of 885 bp in length  
 8037 gap of unknown length  
 8948: contig of 912 bp in length  
 8949 gap of unknown length  
 9836: contig of 888 bp in length  
 10705: contig of 869 bp in length  
 11579: gap of unknown length  
 12459: contig of 880 bp in length  
 11580 gap of unknown length  
 13321: contig of 862 bp in length  
 12460 gap of unknown length  
 13322 gap of unknown length  
 14190: contig of 869 bp in length  
 14191 gap of unknown length  
 15058: contig of 868 bp in length  
 15059 gap of unknown length  
 15923: contig of 862 bp in length  
 15924 gap of unknown length  
 16802: contig of 879 bp in length  
 16803 gap of unknown length  
 17680: contig of 878 bp in length  
 17681 gap of unknown length  
 18548: contig of 868 bp in length  
 19413: gap of unknown length  
 20289: contig of 876 bp in length  
 19414 gap of unknown length  
 21156: contig of 867 bp in length  
 20290 gap of unknown length  
 22031: contig of 875 bp in length  
 21157 gap of unknown length  
 22916: contig of 885 bp in length  
 22032 gap of unknown length  
 22917 gap of unknown length  
 23786: contig of 870 bp in length  
 23787 gap of unknown length  
 24655: contig of 869 bp in length  
 24656 gap of unknown length  
 25519: contig of 864 bp in length  
 25520 gap of unknown length  
 26583: contig of 864 bp in length

26384 gap of unknown length  
 27253: contig of 870 bp in length  
 27254 gap of unknown length  
 28119: contig of 866 bp in length  
 28120 gap of unknown length  
 28987: contig of 869 bp in length  
 28988 gap of unknown length  
 29861: contig of 874 bp in length  
 29862 gap of unknown length  
 30737: contig of 876 bp in length  
 30738 gap of unknown length  
 31601: contig of 864 bp in length  
 31602 gap of unknown length  
 32481: contig of 880 bp in length  
 32482 gap of unknown length  
 33447: contig of 866 bp in length  
 33448 gap of unknown length  
 34218: contig of 871 bp in length  
 34219 gap of unknown length  
 35081: contig of 863 bp in length  
 35082 gap of unknown length  
 35948: contig of 867 bp in length  
 35949 gap of unknown length  
 36814: contig of 866 bp in length  
 36815 gap of unknown length  
 37687: contig of 873 bp in length  
 37688 gap of unknown length  
 38583: contig of 896 bp in length  
 38584 gap of unknown length  
 39452: gap of unknown length  
 39453 contig of 882 bp in length  
 40334: contig of 882 bp in length  
 40335 gap of unknown length  
 41201: contig of 867 bp in length  
 41202 gap of unknown length  
 42067: contig of 866 bp in length  
 42068 gap of unknown length  
 42933: gap of unknown length  
 42934 contig of 868 bp in length  
 43801: contig of 868 bp in length  
 43802 gap of unknown length  
 44675: contig of 874 bp in length  
 44676 gap of unknown length  
 45538: contig of 863 bp in length  
 45539 gap of unknown length  
 46405: contig of 867 bp in length  
 46406 gap of unknown length  
 47269: contig of 864 bp in length  
 47270 gap of unknown length  
 48133: contig of 870 bp in length  
 48134 gap of unknown length  
 48999: contig of 860 bp in length  
 48140 gap of unknown length  
 49000 gap of unknown length  
 49866: contig of 867 bp in length  
 49867 gap of unknown length  
 50729: contig of 863 bp in length  
 50730 gap of unknown length  
 51588: contig of 859 bp in length  
 51589 gap of unknown length  
 52460: contig of 872 bp in length  
 52461 gap of unknown length  
 53323: contig of 863 bp in length  
 53324 gap of unknown length  
 54189: contig of 867 bp in length  
 54190 gap of unknown length  
 55053: contig of 864 bp in length  
 55054 gap of unknown length  
 55910: contig of 857 bp in length  
 55911 gap of unknown length  
 56776: contig of 866 bp in length  
 56777 gap of unknown length  
 57665: contig of 893 bp in length

57670 58529: contig of 860 bp in length  
 58530 59404: gap of unknown length  
 59405 60263: contig of 875 bp in length  
 60264 61140: gap of unknown length  
 61141 61997: contig of 857 bp in length  
 61998 62885: contig of 888 bp in length  
 62886 63756: gap of unknown length  
 63757 64624: contig of 868 bp in length  
 64625 65487: gap of unknown length  
 65488 66346: contig of 859 bp in length  
 66347 67207: gap of unknown length  
 67207 68347: contig of 861 bp in length

Query Match  
 Best Local Similarity 100.0%; Pred. No. 8.8e-05;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2660 aaatatacagaaagagctgaaagagagaa 2691  
 Db 31189 AANTATACGAAAGAGTTGGAATGACGA 31158

RESULT 17  
 LOCUS AC022466 83775 bp DNA RTG 04-FEB-2000  
 DEFINITION Homo sapiens chromosome 15 clone CTD-3027K24 map 15q23. LOW-PASS  
 SEQUENCE SAMPLING.  
 ACCESSION AC022466.1 GI:6899623  
 VERSION AC022466.1  
 KEYWORDS Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 83775)  
 Rowen L., Madan A., Qin S., Abassi N., Baradaran L., Birditt B.,  
 Rowen L., Dots M., Dickhoff R., Fleetwood P., Harrison G.,  
 James R., Kaur A., Madan A., Owen K.P., Ratcliffe A., Shaffer T.,  
 and Hood L.  
 Sequencing of human chromosome 15 D15S146-D15S117 region  
 Unpublished  
 2 (bases 1 to 83775)  
 Rowen L., Madan A., Qin S., Abassi N., Baradaran L., Birditt B.,  
 Rowen L., Dots M., Dickhoff R., Fleetwood P., Harrison G.,  
 James R., Kaur A., Madan A., Owen K.P., Ratcliffe A., Shaffer T.,  
 and Hood L.  
 Submitted (04-FEB-2000) Multimegabase Sequencing Center, University  
 of Washington, PO BOX 357730, Seattle, WA 98195, USA  
 TITLE JOURNAL  
 REFERENCE AUTHORS  
 COMMENT

Center: Genome Center  
 Center: Multimegabase Sequencing Center  
 Web site: [http://chroma.mbt.washington.edu/mag\\_v99](http://chroma.mbt.washington.edu/mag_v99)  
 Contact: gearman@u.washington.edu  
 Sequencing vector: pUC18: 108752  
 Chemistry: Dye-terminator Big Dye; 10% of reads  
 Assembly program: Phrap; version 0.990399

\* NOTE: This record contains 96 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N's are used to indicate the reads  
 \* and the order in which they appear is completely

arbitrary. Low-pass sequence sampling is useful for  
 identifying clones that may be gene-rich and allows  
 overlap relationships among clones to be deduced.  
 However, it should not be assumed that this clone  
 will be sequenced to completion. In the event that  
 the record is updated, the accession number will  
 be preserved.

887: contig of 887 bp in length  
 888 1780: gap of unknown length  
 1781 2645: contig of 865 bp in length  
 2646 3525: contig of 880 bp in length  
 3526 4401: contig of 876 bp in length  
 4402 5290: gap of unknown length  
 5291 6155: contig of 865 bp in length  
 6156 7151: contig of 996 bp in length  
 7152 8036: contig of 885 bp in length  
 8037 8948: gap of unknown length  
 8949 9836: contig of 888 bp in length  
 9837 10705: contig of 869 bp in length  
 10706 11579: gap of unknown length  
 11580 12459: gap of unknown length  
 12460 13322: contig of 862 bp in length  
 13322 14190: contig of 869 bp in length  
 14191 15058: gap of unknown length  
 15059 15923: gap of unknown length  
 15924 16802: contig of 879 bp in length  
 16803 17680: contig of unknown length  
 17681 18548: contig of 868 bp in length  
 18549 19413: contig of 866 bp in length  
 19414 20289: contig of 876 bp in length  
 20290 21156: gap of unknown length  
 21157 22031: gap of unknown length  
 22032 22916: gap of unknown length  
 22917 23786: gap of unknown length  
 23787 24655: contig of 869 bp in length  
 24656 25519: gap of unknown length  
 25520 26383: gap of unknown length  
 26384 27253: gap of unknown length  
 27254 28119: gap of unknown length  
 28120 28987: contig of 866 bp in length

```

28988      29861: contig of 874 bp in length
      29862      30737: contig of 876 bp in length
      30738      31601: gap of unknown length
      31602      32481: contig of 864 bp in length
      32482      33347: gap of unknown length
      33348      34218: contig of 871 bp in length
      34219      35081: gap of unknown length
      35082      35949: gap of unknown length
      35949      36815: contig of 866 bp in length
      36815      37687: gap of unknown length
      37688      38583: contig of 873 bp in length
      38584      39452: gap of unknown length
      39453      40334: contig of 863 bp in length
      40335      41201: gap of unknown length
      41202      42067: contig of 867 bp in length
      42068      42933: gap of unknown length
      42934      43801: contig of 866 bp in length
      43802      44675: gap of unknown length
      44676      45538: contig of 874 bp in length
      45539      46405: gap of unknown length
      46406      47270: contig of 863 bp in length
      47271      48139: gap of unknown length
      48140      48993: contig of 867 bp in length
      48994      49866: gap of unknown length
      49867      50729: contig of 867 bp in length
      50730      51588: gap of unknown length
      51589      52460: contig of 859 bp in length
      52461      53323: gap of unknown length
      53324      54190: contig of 872 bp in length
      54191      55053: gap of unknown length
      55054      55910: contig of 863 bp in length
      55911      56776: gap of unknown length
      56777      57659: contig of 866 bp in length
      57660      58529: gap of unknown length
      58530      59404: contig of 863 bp in length
      59405      60263: gap of unknown length
      60264      61140: contig of 889 bp in length
      61141      61997: gap of unknown length
      61998      62865: contig of 857 bp in length
      62866      63756: gap of unknown length
      63757      64624: contig of 868 bp in length
      64625      65487: gap of unknown length
      65488      66346: contig of 863 bp in length
      66347      67207: gap of unknown length
      67208      68061: contig of 861 bp in length
      68062      68918: gap of unknown length
      68919      69785: contig of 866 bp in length
      69786      70653: gap of unknown length
      70654      71520: contig of 871 bp in length
      71521      72388: gap of unknown length
      72389      73255: contig of 866 bp in length
      73256      74123: gap of unknown length
      74124      75000: contig of 863 bp in length
      75001      75868: gap of unknown length
      75869      76735: contig of 866 bp in length
      76736      77603: gap of unknown length
      77604      78471: contig of 863 bp in length
      78472      79339: gap of unknown length
      79340      80207: contig of 866 bp in length
      80208      81075: gap of unknown length
      81076      81943: contig of 863 bp in length
      81944      82811: gap of unknown length
      82812      83679: contig of 866 bp in length
      83680      84548: gap of unknown length
      84549      85416: contig of 863 bp in length
      85417      86284: gap of unknown length
      86285      87152: contig of 866 bp in length
      87153      88020: gap of unknown length
      88021      88888: contig of 863 bp in length
      88889      89756: gap of unknown length
      89757      90624: contig of 866 bp in length
      90625      91492: gap of unknown length
      91493      92360: contig of 863 bp in length
      92361      93228: gap of unknown length
      93229      94096: contig of 866 bp in length
      94097      94965: gap of unknown length
      94966      95833: contig of 863 bp in length
      95834      96701: gap of unknown length
      96702      97569: contig of 866 bp in length
      97570      98438: gap of unknown length
      98439      99306: contig of 863 bp in length
      99307      100000: gap of unknown length

```







SOURCE human.  
ORGANISM Homo sapiens  
Phylum: Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 91552)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
2 (bases 1 to 91552)  
REFERENCE DOE Joint Genome Institute.  
AUTHORS Direct Submission (2001) Production Sequencing Facility, DOE Joint  
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL On Jul 18, 2000 this sequence version replaced gi:7708979.  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Project Name: 104792, H314  
Center Project Name: CTR78SK6\_10514  
-----  
Summary Statistics  
Consensus quality: 82080 bases at least Q40  
Consensus length: 90076 bases at least Q20  
Consensus quality: 86633 bases at least Q20  
Estimated insert size: 112000; pulse field gel estimation  
Estimated insert size: 112000; pulse field gel estimation  
Estimated coverage: 8.82 in Q20 bases; sum-of-coverage estimation  
Quality coverage: 8.36 in Q20 bases; sum-of-coverage estimation.  
NOTE: This is a 'working draft' sequence. It currently  
consists of 10 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the pieces  
of the gaps between them are based on estimates that have  
been provided by the Illumina platform.  
\* The contig (112000) placed  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\*  
1 15685: contig of 15685 bp in length  
15686 15785: gap of unknown length  
15786 22694: contig of 6909 bp in length  
22695 22794: gap of unknown length  
22795 26047: contig of 3753 bp in length  
26048 26145: gap of unknown length  
26146 41115: contig of 15068 bp in length  
41115 41316: gap of unknown length  
41316 70247: contig of 28932 bp in length  
70248 70348: gap of unknown length  
70348 79105: contig of 8758 bp in length  
79106 79206: gap of unknown length  
79206 83242: contig of 4037 bp in length  
83243 83343: gap of unknown length  
83343 86843: contig of 3500 bp in length  
86843 86946: gap of unknown length  
86946 89101: contig of 2155 bp in length  
89101 89201: gap of unknown length  
89201 91552: contig of 2352 bp in length.  
Location/Qualifiers  
1. 91552  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CCTB-10514"  
/clone\_lib="Cattrech human BAC library B"  
BASE COUNT 29019 a 17406 c 16089 g 28138 t 900 others  
ORIGIN

Query Match 0.5% Score 24; DB 2; Length 91552;  
Best Local Similarity 100.0%; Pos 1; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0;

cy 4707 atccatcaaaaaaaaaaaaaaa 4730  
|||||  
Db 72261 ATCCATCAAAAAAAAAAAAAA 72284  
-----  
RESULT 22  
AC091820/c 104749 bp DNA HTG 09-JUN-2001  
104749 bp DNA HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-286N12, WORKING DRAFT SEQUENCE,  
8 unordered pieces.  
ACCESSION AC091820  
VERSION AC091820.1 GI:14333756  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEPIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Phylum: Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 104749)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
2 (bases 1 to 104749)  
REFERENCE DOE Joint Genome Institute.  
AUTHORS Direct Submission (2001) Production Sequencing Facility, DOE Joint  
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Project Name: 285327  
Center Project Name: CTR78SKC\_286N12  
Center Project Name: CTR78SKC\_286N12  
-----  
Summary Statistics  
Consensus quality: 94051 bases at least Q40  
Consensus length: 100633 bases at least Q20  
Consensus quality: 101439 bases at least Q20  
Estimated insert size: 130000; pulse field gel estimation  
Estimated insert size: 130000; pulse field gel estimation  
Estimated coverage: 9.43 in Q20 bases; sum-of-coverage estimation  
Quality coverage: 9.43 in Q20 bases; sum-of-coverage estimation.  
NOTE: This is a 'working draft' sequence. It currently  
consists of 8 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1447: contig of 1347 bp in length  
1448 1448: gap of unknown length  
1448 3345: contig of 1798 bp in length  
3345 3346: gap of unknown length  
3346 3346: contig of 2109 bp in length  
3346 5554: gap of unknown length  
5554 5555: gap of unknown length  
5555 12539: contig of 6983 bp in length  
12539 12540: gap of unknown length  
12540 22038: gap of unknown length  
22038 22039: gap of unknown length  
22039 46175: contig of 24037 bp in length  
46175 46276: gap of unknown length  
46276 69407: contig of 23132 bp in length  
69407 69408: gap of unknown length  
69408 104749: contig of 35242 bp in length.  
Location/Qualifiers  
1. 104749  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-286N12"

FEATURES  
source





```

repeatL_region /note=match: GSS: Em:AQ018957*
23827..23908
/note=41 copies 2 mer 1t 68% conserved*
repeatL_region repeatL_region
23828..23923
/note=8 copies 12 mer 68% conserved*
23944..24190
/note=12 copies 12 mer 68% conserved*
24226..24423 repeat: matches 265..521 of consensus*
/note=17R33 repeat: matches 3..200 of consensus*
24610..24744
/note=45 copies 3 mer 59% conserved*
24644..24739
/note=8 copies 12 mer 69% conserved*
24740..24859
/note=12 copies 12 mer 70% conserved*
24746..24859
/note=38 copies 3 mer 67% conserved*
25323..25496
/note=MIR repeat: matches 56..246 of consensus*
25604..25756
/note=12 repeat: matches 2333..2491 of consensus*
25891..26193
/note=AluSg repeat: matches 1..303 of consensus*
27141..27213
/note=MIR repeat: matches 81..148 of consensus*
27443..27722
/note=AluSg repeat: matches 1..281 of consensus*
/note=match(27675..27862)
/note=match: GSS: Em:AQ020767*
28050..28097
/note=4 copies 12 mer 83% conserved*
28118..28197
/note=MIR repeat: matches 1828..2656 of consensus*
29054..29346
/note=AluSg repeat: matches 1..291 of consensus*
29729..29793
/note=12 repeat: matches 2622..2693 of consensus*
29836..29915
/note=12 repeat: matches 2593..2680 of consensus*
29916..29987
/note=12 repeat: matches 2 mer at 66% conserved*
30841..31048
/note=MIR repeat: matches 20..226 of consensus*
31386..31420
/note=MIR repeat: matches 1..35 of consensus*
31715..32085
/note=12 repeat: matches 5803..6168 of consensus*
32844..33672
/note=12 repeat: matches 5319..6157 of consensus*
33676..33790
/note=12 repeat: matches 5319..5434 of consensus*
/note=12 repeat: matches 5319..5434 of consensus*
Query Match 0.5%: score 24; DB 9; length 113599;
Best Local Similarity 100.0%: Pred. No. 1.4;
Matches 24: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OR 4707 altcatcaaaaaaaaaaaaaa 4730
DB 100557 ATCCATCAAAAAAAAAAAAAA 100580

RESULT 25
AF165138 127999 bp DNA PRI 16-JUL-1999
LOCUS Homo sapiens PAC B99112 derived from chromosome 21p11.2, complete
ACCESSION AF165138
VERSION AF165138.1 GI:5499744
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

---

```

REFERENCE 1 (bases 1 to 127999)
AUTHORS Brieschmidt,K., Nordtsch.G., Groet,J., Nizetic,D., Hildmann,T.,
Drescher,B., Weber,J., Menzel,U., Schaltevoy,R., Vaspo,M.-L. and
Grotzer,T.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..127999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/ncbi_taxid="21"
/map22="PAC B99112"
383..698
/rpt_family="AluSg"
/evidence=not_experimental
complement(1155..1650)
/rpt_family="MIR1D"
/evidence=not_experimental
complement(4924..5085)
/rpt_family="MIR2"
/evidence=not_experimental
3491..3511
/rpt_family="Alu"
/evidence=not_experimental
3512..3728
/rpt_family="AluSg"
/evidence=not_experimental
3729..3832
/rpt_family="Alu"
/evidence=not_experimental
complement(3815..4015)
/rpt_family="MIR3"
/evidence=not_experimental
complement(4350..4457)
/rpt_family="AluSg/X"
/evidence=not_experimental
complement(4924..5085)
/note="MZEF, score = 86.8%"
5820..6128
/evidence=not_experimental
/rpt_family="AluX"
/evidence=not_experimental
complement(7647..7835)
/note="McKenzie, score = 2.77%, comment = 'terminal'exon 189
b frame 5820"
/evidence=not_experimental
complement(7647..7835)
/note="GRAIL, score = 49.000%, comment = 'marginal shadow'
/evidence=not_experimental
7651..7841
/note="GRAIL, score = 76.000%, comment = 'excellent'
/evidence=not_experimental
778..7841
/note="Xpound exon prediction, score = 78% (4%)
/evidence=not_experimental
complement(8675..9638)
/rpt_family="L2"
/evidence=not_experimental
9364..9590
/note="GRAIL, score = 52.000%, comment = 'good shadow'
complement(1035..1078)
/note="homology = 74.30%, score = 20, counts = 5"
/rpt_type="tandem"
/evidence=not_experimental
complement(10449..10643)
/note="GRAIL, score = 100.000%, comment = 'excellent'
/evidence=not_experimental
complement(10477..10643)
/note="GRAIL, score = 10.66%, comment = 'Internal_exon

```

167 bp frame: 0 phase: 2\*  
 /evidence-not\_experimental  
 complement(10477..10643)  
 /note="MZEF, score = 87.18"  
 /evidence-not\_experimental  
 complement(11473..11546)  
 /note="GRAIL, score = 75.000%, comment = excellent"  
 /evidence-not\_experimental  
 complement(13842..14122)  
 /note="MZEF, score = 53.58"  
 /evidence-not\_experimental  
 complement(15687..15797)  
 /rpt\_family="MER34"  
 /evidence-not\_experimental  
 complement(15798..15936)  
 /rpt\_family="FLAM.C"  
 /evidence-not\_experimental  
 complement(15937..16353)  
 /note="GRAIL, score = 73.217%  
 /evidence-not\_experimental  
 complement(17570..17672)  
 /note="Xpound exon prediction, score = 79% (2%)"  
 /evidence-not\_experimental  
 18537..19236  
 /rpt\_family="L1PA7"  
 /evidence-not\_experimental  
 1854..19081, score = 43.000%, comment = marginal shadow"  
 /evidence-not\_experimental  
 complement(20386..20491)  
 /note="GRAIL, score = 93.000%, comment = excellent"  
 /evidence-not\_experimental  
 complement(21159..21288)  
 /note="MZEF, score = 97%"  
 /evidence-not\_experimental  
 2128..21795  
 /rpt\_family="M1P8"  
 /evidence-not\_experimental  
 21843..22125  
 /rpt\_family="AlusX"  
 /evidence-not\_experimental  
 complement(22971..23121)  
 /note="GenScan, score = 20.21%, comment = Internal-exon  
 151 bp frame: 0 phase: 1\*  
 /evidence-not\_experimental  
 complement(23571..23710)  
 /note="GRAIL, score = 100.000%, comment = excellent"  
 /evidence-not\_experimental  
 complement(25058..25264)  
 /note="GenScan, score = 16.35%, comment = Internal-exon  
 207 bp frame: 1 phase: 0\*  
 /evidence-not\_experimental  
 complement(25058..25264)  
 /note="GRAIL, score = 99.000%, comment = excellent"  
 /evidence-not\_experimental  
 complement(25058..25264)  
 /note="MZEF, score = 91.84"  
 /evidence-not\_experimental  
 25419..25679  
 /rpt\_family="AluJ0"  
 /evidence-not\_experimental  
 complement(25544..26036)  
 /rpt\_family="M1P8"  
 /evidence-not\_experimental  
 26953..27141  
 /note="MZEF, score = 70.54"  
 /evidence-not\_experimental  
 27852..28117  
 /rpt\_family="M1P8"  
 /evidence-not\_experimental

exon 28226..28275  
 /note="GRAIL, score = 41.000%, comment = marginal"  
 /evidence-not\_experimental  
 repeat\_region 28226..28433  
 /rpt\_family="L1P8a"  
 /evidence-not\_experimental  
 repeat\_region 28410..33376  
 /rpt\_family="L1P8a"  
 /evidence-not\_experimental  
 exon 30879..31172  
 /note="GRAIL, score = 40.000%, comment = marginal"  
 /evidence-not\_experimental  
 repeat\_region 32575..32737  
 /note="GRAIL, score = 53.000%, comment = good"  
 /evidence-not\_experimental  
 repeat\_region 33705..34074  
 /rpt\_family="AlusC"  
 /evidence-not\_experimental  
 exon 33880..33935  
 /note="GRAIL, score = 42.000%, comment = marginal"  
 /evidence-not\_experimental  
 exon 34382..34485  
 /note="MZEF, score = 56.54"  
 /evidence-not\_experimental  
 repeat\_region 34387..34693  
 /rpt\_family="AluY"  
 /evidence-not\_experimental  
 exon 34693..34855  
 /note="GenScan, score = 18.66%, comment = Internal-exon  
 152 bp frame: 2 phase: 2\*  
 /evidence-not\_experimental  
 complement(34839..34990)  
 /note="GRAIL, score = 70.000%, comment = good"  
 /evidence-not\_experimental  
 complement(34839..34990)  
 Query Match 0.5% Score 24: DB 9: Length 127999;  
 Database: Genbank; File No: 144;  
 Matches 24: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4707 attccatcaaaaaaaaaaaaaa 4730  
 DB 103203 ATCCATCAAAAAAAAAAAAAA 103226  
 RESULT 26  
 AC007992 146842 bp DNA HTG 09-AUG-2001  
 Homo sapiens clone RP11-44N17. \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION  
 ordered pieces.  
 AC007992 AC007992.9 GI:14150923  
 VERSION  
 KEYWORDS HTG; HTGS; PULFOP; HTGS; ACTIVEPIN.  
 ORGANISM  
 human.  
 Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 146842)  
 TITLE  
 Homo sapiens, clone RP11-44N17  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 146842)  
 Birren,B., Linton,J., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A.,

AL359743/C







REFERENCE  
AUTHORS

2 (bases 1 to 156864)

TITLE  
JOURNAL  
COMMENT

2 (bases 1 to 156864)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldvini,J., Barne,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., Deatellano,K., Dewar,K., Donlan,L., Doyle,M.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
 Gargan,J., Gargay,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
 Johnson,J., Johnson,S., Locke,K., MacDonald,P., Marquis,N., Kiehn,J.,  
 McMan,P., McKus,A., McKernan,K., McLaughlin,J., McElrath,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Telamas,J.,  
 Testaye,S., Tirelli,A., Vasilev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
 Direct Submission (1998-1999), Whitehead Institute/MIT Center for Genome  
 Research, 700 State Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www.seq.wi.mit.edu

Contact: sequenc@genomics.genome.wi.mit.edu  
 Center project name: L2629

Center clone name: 399.L.5

Summary Statistics

Sequencing vector: M13; M7815; 47% of reads  
 Sequencing vector: plasmid; n/a; 53% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.96031  
 Consensus quality: 15132 bases at least Q30  
 Consensus quality: 15576 bases at least Q20  
 Insert size: 151000; agarose-efp  
 Insert size: 155964; sum-of-contigs  
 Quality coverage: 10.6 in Q20 bases; agarose-efp

NOTE: This is a 'working draft' sequence. It currently  
 consists of 10 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. The gaps between contigs are not known  
 runs of N's, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 41811: contig of 41811 bp in length  
 41812 41911: gap of 100 bp  
 41912 43011: contig of 1100 bp in length  
 43011 43111: gap of 100 bp  
 43111 44111: gap of 100 bp in length  
 44111 44557: contig of 1141 bp in length  
 44557 45557: gap of 100 bp  
 45557 53893: contig of 8236 bp in length  
 53893 53993: gap of 100 bp  
 53993 63288: contig of 9295 bp in length  
 63288 63388: gap of 100 bp  
 63388 79429: contig of 15932 bp in length  
 79429 116018: gap of 100 bp  
 116018 116118: gap of 100 bp  
 116118 147940: contig of 31822 bp in length  
 147940 148041: gap of 100 bp  
 148041 156864: contig of 8824 bp in length.

FEATURES  
SOURCE

1. 156864  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

/chromosome="18"  
 /map="18"  
 /clone="RP11-299L5"  
 /clone\_id="RP11-299L5 Human Male BAC"  
 1. 41811  
 /note="assembly-fragment"  
 /note="slide:left"  
 41912-43011  
 /note="assembly-fragment"  
 43112-44316  
 /note="assembly-fragment"  
 44417-45557  
 /note="assembly-fragment"  
 45558-53893  
 /note="assembly-fragment"  
 53894-53928  
 /note="assembly-fragment"  
 63389-79320  
 /note="assembly-fragment"  
 79421-116018  
 /note="assembly-fragment"  
 116119-147940  
 /note="assembly-fragment"  
 148041-156864  
 /note="assembly-fragment"  
 /note="slide:right"

BASE COUNT 46129 a 32086 c 30930 g 46819 t 900 others

## ORIGIN

Query Match 0.5% Score 24: DB 2: Length 156864;  
 Best Local Similarity 100.0% Pred. No. 1.4;  
 Matches 24: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 0y 4707 atccatcaaaaaaaaaaaaaa 4730  
 DB 133293 ATCCATCAAAAAAAAAAAAA 133316

RESULT 31  
 LOCUS AC02597/c 156523 bp DNA 09-MAY-2001  
 DEFINITION Homo sapiens clone RP11-296L5 map 1, WORKING DRAFT  
 SEQUENCE, 9 unordered pieces.  
 AC02597.5 GI:1376956  
 VERSION  
 KEYWORDS  
 HGTG: HGTG-PHASE1; HGTG-DRAFT; HGTG-FULLTOP.  
 SOURCE  
 Human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

(bases 1 to 156523) Nusbaum,C. and Lander,E.

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

2 (bases 1 to 156864)



Center clone name: 24.L121  
 Assembly Statistics  
 Sequencing vector: M13: W7815: 100% of reads  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.960731  
 Consensus quality: 145782 bases at least Q40  
 Consensus quality: 152368 bases at least Q30  
 Consensus quality: 154985 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Insert size: 157571; sum-of-contigs  
 Quality coverage: 3.4 in Q20 bases; sum-of-contigs  
 Quality coverage: 3.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. 1205: contig of 1205 bp in length

1206 1305: gap of 100 bp  
 1306 2709: contig of 1404 bp in length  
 2710 2809: gap of 100 bp  
 2810 3985: contig of 1176 bp in length  
 3986 4085: gap of 100 bp  
 4086 6516: contig of 2531 bp in length  
 6517 6807: gap of 100 bp  
 6808 9107: gap of 100 bp in length  
 9108 11295: contig of 2188 bp in length  
 11296 11395: gap of 100 bp  
 11396 14811: contig of 3416 bp in length  
 14812 14911: gap of 100 bp  
 14912 18366: contig of 3453 bp in length  
 18367 18492: gap of 100 bp  
 18493 22123: contig of 3656 bp in length  
 22123 22222: gap of 100 bp  
 22223 22546: contig of 3324 bp in length  
 22547 25646: gap of 100 bp  
 25647 29226: contig of 3580 bp in length  
 29227 29326: gap of 100 bp  
 29327 32285: contig of 2859 bp in length  
 32286 33553: gap of 100 bp  
 33554 36206: contig of 3721 bp in length  
 36207 40193: contig of 3987 bp in length  
 40194 40293: gap of 100 bp  
 40294 45635: contig of 5342 bp in length  
 45636 45735: gap of 100 bp  
 45736 50445: contig of 4710 bp in length  
 50446 50545: gap of 100 bp  
 50546 54232: contig of 3721 bp in length  
 54233 60095: contig of 5863 bp in length  
 60096 60195: gap of 100 bp  
 60196 64607: contig of 4412 bp in length  
 64608 64707: gap of 100 bp  
 64708 71939: contig of 7232 bp in length  
 71940 72039: gap of 100 bp  
 72040 77453: contig of 5323 bp in length  
 77454 82978: contig of 5523 bp in length  
 82979 83078: gap of 100 bp  
 83079 89653: contig of 6575 bp in length  
 89654 89753: gap of 100 bp  
 89754 96604: contig of 6851 bp in length  
 96605 96704: gap of 100 bp  
 96705 103170: contig of 6666 bp in length  
 103171 112353: contig of 9083 bp in length  
 112354 112453: gap of 100 bp

FEATURES

source

112454 121857: contig of 9404 bp in length  
 \* 121858 121957: gap of 100 bp  
 \* 121958 129572: contig of 7615 bp in length  
 \* 129573 129672: gap of 100 bp  
 \* 129673 137272: contig of 7600 bp in length  
 \* 137273 137372: gap of 100 bp  
 \* 137373 160270: contig of 22899 bp in length.  
 local:0270:contigfiles  
 1 160271  
 /orient:+"Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_rpt1.24121  
 /clone\_1lb="RPI-11 Human Male BAC"  
 1. 1205  
 /note="assembly-fragment"  
 1306 2709  
 /note="assembly-fragment"  
 2810 3985  
 /note="assembly-fragment"  
 clone\_end:17  
 vector\_side:left"  
 4086 6416  
 /note="assembly-fragment"  
 6517 9007  
 /note="assembly-fragment"  
 9008 11295  
 /note="assembly-fragment"  
 11396 14811  
 /note="assembly-fragment"  
 14912 18366  
 /note="assembly-fragment"  
 18467 22122  
 /note="assembly-fragment"  
 22123 22546  
 /note="assembly-fragment"  
 25647 29226  
 /note="assembly-fragment"  
 29327 32285  
 /note="assembly-fragment"  
 32286 33553  
 /note="assembly-fragment"  
 33554 36106  
 /note="assembly-fragment"  
 36207 40294  
 /note="assembly-fragment"  
 40294 45635  
 /note="assembly-fragment"  
 45636 45735  
 /note="assembly-fragment"  
 45736 50445  
 /note="assembly-fragment"  
 50446 54132  
 /note="assembly-fragment"  
 54233 60095  
 /note="assembly-fragment"  
 60096 64607  
 /note="assembly-fragment"  
 64708 71939  
 /note="assembly-fragment"  
 72040 77355  
 /note="assembly-fragment"  
 clone\_end:576  
 vector\_side:left"  
 77454 82978  
 /note="assembly-fragment"  
 83079 89653  
 /note="assembly-fragment"  
 89754 96604  
 /note="assembly-fragment"  
 96705 103170  
 /note="assembly-fragment"  
 103171 112353  
 /note="assembly-fragment"  
 112454 121857  
 /note="assembly-fragment"  
 121958 129572  
 /note="assembly-fragment"









```

/notes="15 copies 6 mer agagag 75% conserved"
repeatL_region 1924..19374
/notes="12 repeat: matches 2539..2704 of consensus"
repeatL_region 19374..19413
/notes="Charile4 repeat: matches 36..76 of consensus"
repeatL_region 19413..19743
/notes="Aluv repeat: matches 1..309 of consensus"
repeatL_region 19744..20098
/notes="Charile4 repeat: matches 76..439 of consensus"
repeatL_region 20739..20786
/notes="12 copies 4 mer caca 87% conserved"
repeatL_region 20786..22083
/notes="23 copies 2 mer ac 89% conserved"
repeatL_region 20752..20787
/notes="6 copies 6 mer acacac 97% conserved"
repeatL_region 20794..21401
/notes="MER82 repeat: matches 52..552 of consensus"
repeatL_region 21409..21709
/notes="MWR28 repeat: matches 3..314 of consensus"
repeatL_region 22304..22382
/notes="MER82 repeat: matches 2215..2293 of consensus"
repeatL_region 22402..22763
/notes="MWR1 repeat: matches 112..371 of consensus"
repeatL_region 23107..23172
/notes="3 copies 22 mer 83% conserved"
repeatL_region 23109..23170
/notes="31 copies 2 mer ac 82% conserved"
repeatL_region 23123..23170
/notes="8 copies 6 mer acacac 89% conserved"
misc_feature 23170..23175
/notes="match: GSS: Em:AD51587"
repeatL_region 24527..24572
/notes="2 copies 23 mer 100% conserved"
repeatL_region 25692..25997
/notes="MER58B repeat: matches 2..341 of consensus"
repeatL_region 27028..27115
/notes="4 copies 22 mer 73% conserved"
repeatL_region 27028..27115
/notes="16 copies 4 mer agag 78% conserved"
repeatL_region 27490..27731
/notes="1LEP4 repeat: matches 5566..5805 of consensus"
repeatL_region 27732..28078
/notes="1LEP4 repeat: matches 5798..6144 of consensus"
repeatL_region 29220..29326
/notes="MER58B repeat: matches 5..128 of consensus"
repeatL_region 29362..29531
/notes="repeat: matches 2245..2417 of consensus"
repeatL_region 29536..29605
/notes="MWR1C repeat: matches 187..465 of consensus"
repeatL_region 29906..30065
/notes="Aluv repeat: matches 124..288 of consensus"
repeatL_region 30066..30327
/notes="Aluv repeat: matches 40..296 of consensus"
repeatL_region 30328..30452
/notes="Aluv repeat: matches 1..124 of consensus"
repeatL_region 30453..30722
/notes="MWR1 repeat: matches 1..187 of consensus"
repeatL_region 30702..30844
/notes="12 repeat: matches 2593..2749 of consensus"
misc_feature 32132..32829
/notes="match: GSS: Em:AD381845"
repeatL_region 32204..32423
/notes="MER58B repeat: matches 1..224 of consensus"
repeatL_region 32991..33229
/notes="repeat: matches 5419..5661 of consensus"
repeatL_region 33206..33507
/notes="1LM4 repeat: matches 3203..5031 of consensus"
repeatL_region 35039..36782
/notes="1LMC repeat: matches 1119..2557 of consensus"
repeatL_region 36819..36956
/notes="1LMC repeat: matches 897..1049 of consensus"
repeatL_region 36971..37156
/notes="MER20 repeat: matches 3..190 of consensus"

```

```

repeatL_region 37997..38159
Query Match 0.5%; Score 24; DB 9; Length 170740;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4707 accatcaaaaaaaaaaaaaa 4730
DB 137501 ATCCATCAAAAAAAAAAAAA 137524

RESULT 37
AC011628/c
LOCUS
DEFINITION
ACCESSION
VERSION
XREF
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.gen.wi.mit.edu
Contact: genome@genomics.wi.mit.edu
Project Information
Center project name: L2100
Center clone name: 399_D_2
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: dye-terminator Big Dye; 30017 of reads
Assembly program: Phred; 52 bases at least Q40
Consensus quality: 169138 bases at least Q30
Consensus quality: 171937 bases at least Q20
Insert size: 18200; agarose-fp
Insert size: 173638; sum-of-continigs
Quality coverage: 5.7 in Q20 bases; sum-of-continigs
Quality coverage: 6.0 in Q20 bases; sum-of-continigs
-----
* NOTE: This is a "working draft" sequence. It currently
* contains the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```











```

FEATURES
  * 76991 119702: contig of 42712 bp in length
  * 119703 119802: gap of unknown length
  * 119803 189594: contig of 66152 bp in length.
Location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="3"
  /clone="RP11-569L16"
  1..1328
misc_feature
  /note="assembly_name:Contig14"
  1429..8721
misc_feature
  /note="assembly_name:Contig17"
  8925..36928
misc_feature
  /note="assembly_name:Contig18"
  vector_side:left"
  vector_side:right"
misc_feature
  /note="assembly_name:Contig19"
  37029..76890
misc_feature
  /note="assembly_name:Contig10"
  76991..119702
misc_feature
  /note="assembly_name:Contig11"
  119803..189594
  vector_side:left"
  vector_side:right"
misc_feature
  /note="assembly_name:Contig11"
  587 others
BASE COUNT 57589 a 35403 c 35563 g 56812 t
ORIGIN
Query Match 0.5% Score 24: DB 2: Length 189594:
Mismatch 100.0% Pred. No. 1.4:
Matches 24: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
Qy 4706 gatccatacaaaaaaaaaaaaaa 4729
Db 39174 GATCCATCAAAAAAAAAAAAAA 39151

RESULT 44
AC023376 189963 bp DNA HTG 04-JUL-2001
LOCUS Homo sapiens chromosome 8 clone RP11-687M23 map 8, WORKING DRAFT
DEFINITION
SEQUENCE 4 unordered pieces.
AC023376
AC023376.4 GI:14595907
HTG: HTGS_MHASE1: HTGS_DRAFT: HTGS_FULFORD:
SOURCE human
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
  1 (bases 1 to 189963)
  Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome 8, Clone RP11-687M23
  Unpublished
  2 (bases 1 to 189963)
  AC023376
  Anderson,S., Baldwin,J., Barnes,N., Bedalov,F., Boguslavsky,L.,
  Bouchalter,B., Brown,A., Burkett,G., Campolongo,A., Castle,A.,
  Choppel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
  Deatellano,K., Dewar,K., Dodge,S., Domio,M., Doyle,M.,
  Galagan,J., Gertzel,P., Fitzhugh,M., Forrest,C., Gage,D.,
  Gelsinger,J., Geyrhofer,S., Glade,S., Goyette,M., Graham,L.,
  Grand-Pierre,N., Grant,G., Hagos,B., Heide,C., Kohn,L.,
  Kelland,J., Landers,T., Larocque,K., Lebecky,J., Levine,R.,
  Lien,C., Liu,C., Locke,K., MacDonald,P., Marguis,N., McCarthy,M.,
  McKean,P., McGurk,A., McKernan,K., McPherson,R., Meltrm,J.,
  Menais,L., Milnova,T., Miranda,C., Mlenaga,V., Morrow,T.,
  Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
  Peterson,K., Pierre,N., Pisanl,C., Polara,V., Raymond,C.,
  Riley,K., Rogov,P., Roldman,D., Roy,A., Santos,R., Savic,N.,
  Severy,P., Spencer,B., Stange,R., Stenhouse,N., Stenhouse,N.,
  Sudramanlan,A., Tolman,J., Tostay,S., Theodore,J., Tiroli,A.,

```

```

TITLE
JOURNAL
COMMENT
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
  Zody,M., Wyman,D., Ye,W.J., Young,G., Zakhon,J., Zimmer,A. and
  others
Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Jul 4, 2001 this sequence version replaced g1.7139677.
  All repeats were identified using RepeatMasker.
  SML: A.F.A. & Green, http://hg19611.mit.edu/hw/RepeatMasker.html
  Center: 96cm Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: M18
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: g1.72093
Center clone name: Stat1b1s
Sequencing vector: Stat1b1s
Sequencing vector: Plasmid: n/a; 53% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 188936 bases at least Q40
Consensus quality: 189476 bases at least Q40
Consensus quality: 189376 bases at least Q20
Insert size: 189600; 189200bp-FP
Insert size: 189663; sum-of-ctnigs
Quality coverage: 10.7 in Q20 bases; agarose-fp
Quality coverage: 10.7 in Q20.
NOTE: This is a 'working draft' sequence. It currently
  consists of 4 contigs. The true order of the pieces and
  is not known and their order in this sequence is identified as
  arbitrary. Gaps between slices of the gaps are unknown.
  This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
  1 37424: contig of 37424 bp in length
  37425 37524: gap of 100 bp
  37525 42149: contig of 4625 bp in length
  42150 42249: gap of 100 bp
  42250 17406: gap of 135057 bp in length
  17407 17406: gap of 100 bp
  17407 189963: contig of 12557 bp in length.
Location/Qualifiers
  1..189963
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="8"
  /map="8:RP11-687M23"
  /clone="RP11-687M23"
  /clone="RP11-687M23"
  1..37424
  /note="assembly_fragment"
  clone_end:576
  /note="assembly_fragment"
  vector_side:left"
  vector_side:right"
misc_feature
  37525..42149
  /note="assembly_fragment"
  42250..42249
  /note="assembly_fragment"
  17407..189963
  /note="assembly_fragment"
  clone_end:17
  vector_side:left"
  vector_side:right"
BASE COUNT 57206 a 36557 c 37255 g 58644 t
ORIGIN
Query Match 0.5% Score 24: DB 2: Length 189963:
Mismatch 100.0% Pred. No. 1.4:
Matches 24: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```





Job time: 15600 sec

```

/misc_feature /chromosome="18"
/misc_feature /note="p11-455M13"
/misc_feature /clone.lib="RXC1-11 Human Male BAC"
1..70
/note="assembly_fragment
clone_end:SP6
vector_side:left"
/misc_feature 171..1131
/note="assembly_fragment"
/misc_feature 1272..1839
/note="assembly_fragment"
/misc_feature 2050..2985
/note="assembly_fragment"
/misc_feature 3086..4155
/note="assembly_fragment"
/misc_feature 4256..5513
/note="assembly_fragment"
/misc_feature 5614..6883
/note="assembly_fragment"
/misc_feature 6984..7525
/note="assembly_fragment"
/misc_feature 7626..9324
/note="assembly_fragment"
/misc_feature 9425..11273
/note="assembly_fragment"
/misc_feature 11374..13899
/note="assembly_fragment"
/misc_feature 14000..15749
/note="assembly_fragment"
/misc_feature 15861..17799
/note="assembly_fragment"
/misc_feature 17900..20571
/note="assembly_fragment"
/misc_feature 20672..23544
/note="assembly_fragment"
/misc_feature 23545..28580
/note="assembly_fragment"
/misc_feature 26681..30718
/note="assembly_fragment"
/misc_feature 30819..37775
/note="assembly_fragment"
/misc_feature 37876..45999
/note="assembly_fragment"
/misc_feature 46100..54092
/note="assembly_fragment"
/misc_feature 54123..61685
/note="assembly_fragment"
/misc_feature 61786..71733
/note="assembly_fragment"
/misc_feature 71834..81685
/note="assembly_fragment"
/misc_feature 81786..89237
/note="assembly_fragment"
/misc_feature 89338..98885
/note="assembly_fragment"
/misc_feature 98886..109757
/note="assembly_fragment"
/misc_feature 109858..118004
/note="assembly_fragment"
/misc_feature 118105..127896

```

Query Match 0.58; Score 24; DB 2; Length 207050;  
 Best Local Similarity 100.0%; Freq. No. 14;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4707 atccatcaaaaaaaaaaaaaaa 4730  
 |||  
 DB 43480 atccatcaaaaaaaaaaaaaaa 43503

Search completed: April 10, 2002, 19:18:41

\_\_\_\_\_



CC sequence of clones c13c.4 (AAVB83132), d422b.6 (AAVB83133), d4410.5  
CC FAD03134, vcr180\_1 (AAVB83135), cr365r.1 (AAVB83136), fh123.5 (AAVB83177),  
CC FAD03134, vcr180\_1 (AAVB83135), cr365r.1 (AAVB83136), fh123.5 (AAVB83177),  
CC ATCC 98415) and the protein has been deposited as accession number AF013900.  
CC biological activities which would make them suitable for treating,  
CC preventing or ameliorating medical conditions in humans and animals  
CC for example, tumour suppression/invasion activity, immune system  
CC boosting activity. The polynucleotides are also believed to be useful  
CC for gene therapy. This polynucleotide was isolated from a human  
CC foetal brain cDNA library.

S0 Sequence 2322 BP; 957 A; 363 C; 498 G; 504 T; 0 other:

Query Match 1.6% Score 74; DB 20; Length 2322;  
Best Local Similarity 100.0%; Pred. NO. 2,5e-19;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4460 gccagacgaccacgaagaatgatgcatttcggcacacacctctttagtgcgcacg 4519  
Db 2054 gccagacgaccacgaagaagaagtatgtcatcttcaggcaaacacctctttagtgcgcacg 2113  
Oy 4520 gttccatcggtgtta 4533  
Db 2114 gggtcactcagtgatga 2127

RESULT 3  
AAC27957  
AAC27957 standard; CDNA; 355 BP.  
AAC27957:  
06-OCT-2000 (first entry)  
Human secreted protein 5' EST, SBO ID NO: 32032.  
Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;  
gene therapy; chromosome mapping; 89.  
Homo sapiens.  
EP103401-A2.  
06-SEP-2000.  
21-FEB-2000; 2000EP-0200610.  
26-FEB-1999; 9905-0122487.  
(GEST ) GENSET.  
Dumas Milne Edwards J., Duclert A., Giordano J.;  
WPI: 2000-500381/45.  
New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
identifying CDNAs and genomic DNAs that correspond to 5' ESTs and for  
diagnostic, forensic, gene therapy and chromosome mapping procedures -  
Claim 1; SBO ID 32032; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from  
mRNAs encoding secreted proteins. No ORF has yet been conclusively  
identified within the present sequence. The 5' ESTs were prepared from  
total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
sequences usually correspond mainly to the 3' untranslated region (UTR)  
of the mRNA because they are often obtained from oligo-dT primed cDNA  
libraries. The 5' ends are not well suited for isolating cDNA sequences  
derived from the 5' ends of mRNAs with intact 5' caps and can therefore no  
longer be used to identify genes. However, the full length cDNAs and longer  
CDNA sequences have been obtained "the full 5' UTR" is conserved among  
5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC In diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

SO Sequence 355 BP; 110 A; 61 C; 62 G; 120 T; 2 other;

Query Match 1.3%; Score 61; DB 21; Length 355;

Best Local Similarity 100.0%; Pred. No. 3.2e-14; Mismatches 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 4460 gacagacagccacgaagaatgcaattatcgcacacaccccttcgtgctgcacag 4519

DB 128 gacagacagccacgaagaatgcaattatcgcacacaccccttcgtgctgcacag 187

Oy 4520 g 4520

DB 188 g 188

#### RESULT 4

AAK41086

AAK41086 standard; cDNA; 356 BP.

18-JUN-1999 (first entry)

Human secreted protein 5' EST SEQ ID NO: 298.

Human; secreted protein; EST; expressed sequence tag; diagnosis;

upstream regulatory sequence; cytokine activity; cell proliferation;

differentiation; haematopoiesis regulation; tissue growth regulation;

reproductive hormone regulation; chemoketic; chemoketic; haemostatic;

thrombolytic; anti-inflammatory; tumour inhibition; ds.

Homo sapiens.

MO9906554-A2.

11-FEB-1999.

31-JUL-1998; 98MO-IB01238.

01-AUG-1997; 97US-0905134.

(GSEST) GENEST.

Duclet A, Dumas Milne Edwards J, Lacroix B;

WPI; 1999-153784/13.

P-PSDB; AAV12253.

New nucleic acids encoding human secreted proteins - obtained from

CNH libraries prepared from kidney, fetal kidney, dystrophic

muscle, muscle and heart tissue

Claim 1: Page 442; 622pp: English.

AAK40826 to AAK41093 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAK41092 and  
 CC AAK41194 to AAK12260, respectively. The proteins given in AAK41092 and  
 CC signal peptide and a secreted protein. The proteins given in AAK41194  
 CC and AAK12260 can be used for producing secreted human gene  
 CC products. They can also be used to develop products for diagnosis and  
 CC therapy. The proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulation  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC and thrombolytic activity, chemoketic/chemoketic activity, haemostatic and  
 CC activity, tumour inhibition activity, anti-inflammatory activity, ds.  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.

CC The sequences can also be used for obtaining corresponding promoter  
 CC regions. The nucleic acids encoding the signal peptide can be used  
 CC for directing extracellular secretion of a polypeptide or the insertion  
 CC of a polypeptide into a membrane, or importing a polypeptide into  
 CC a cell.

SO Sequence 356 BP; 110 A; 61 C; 63 G; 120 T; 2 other;

Query Match 1.3%; Score 61; DB 20; Length 356;

Best Local Similarity 100.0%; Pred. No. 3.2e-14; Mismatches 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 4460 gacagacagccacgaagaatgcaattatcgcacacaccccttcgtgctgcacag 4519

DB 128 gacagacagccacgaagaatgcaattatcgcacacaccccttcgtgctgcacag 187

Oy 4520 g 4520

DB 188 g 188

#### RESULT 5

AAK40197

AAK40197 standard; DNA; 1980 BP.

02-JUL-1999 (first entry)

Sequence of C3Vs gene.

Cancer associated antigen; diagnosis; research; treatment; human;

breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

prostate cancer; ss.

Homo sapiens.

MO9904265-A2.

28-JAN-1999.

15-JUL-1998; 98MO-US14679.

22-JUN-1998; 98US-0102322.

17-JUL-1997; 97US-0896164.

10-OCT-1997; 97US-0061599.

10-OCT-1997; 97US-0061765.

11-OCT-1997; 97US-0948705.

11-OCT-1997; 97GB-0021697.

(UDDM-) UDDMIG INST CANCER RES.

Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

Pfreundschuh M, Sahlin U, Scanlan MJ, Stockert E;

Tureci O;

WPI; 1999-132448/11.

New isolated cancer associated nucleic acids and polypeptides -

isolated using sera from cancer patients, used to develop products

for the diagnosis, monitoring or treatment of cancers

Claim 67: Page 776-777; 787pp: English.

The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer-associated protein, comprising: (a) contacting  
 CC a nucleic acid isolated from a subject with an agent that specifically  
 CC binds to the NMM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NMM or the expression product as a

CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterized by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC that are cancer associated antigen precursors expressed in human breast  
 CC cancer, lung cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.

XX Sequence 1980 BP; 683 A; 351 C; 511 G; 435 T; 0 other:

Query Match 1.0%; Score 45; DB 20; Length 1980;  
 DB Local Similarity 100.0%; Pred. No. 4, 2e-08;  
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 695 tctctctcaaacatctctcagcagatctgctccacgagatg 740  
 Db 350 tctctctcaaacatctctcagcagatctgctccacgagatg 394

RESULT 6

AAAC27789  
 ID AAC27789 standard; CDNA: 361 BP.

AC AAC27789;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 31864.

XX Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;  
 XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dimos Milne Edwards J, Duclet A, Clordano J;

XX WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS Claim 1: SEQ ID 31864; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 3' ESTs were prepared from  
 CC total human RNAs or poly(A) RNAs derived from 3' untranslated  
 CC sequences usually correspond mainly to the 3' untranslated  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC 5' ESTs have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs have been used for isolating cDNAs, 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 361 BP; 112 A; 75 C; 89 G; 83 T; 2 other;

Query Match 0.8%; Score 37; DB 21; Length 361;  
 DB Local Similarity 100.0%; Pred. No. 6, 3e-05;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 885 gtaggcacacatagtcacgtcgtatagatgag 921  
 Db 197 gtaggcacacatagtcacgtcgtatagatgag 233

RESULT 7

AAAF6083  
 ID AAFF6083 standard; CDNA: 371 BP.

AC AAFF6083;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 1839.

XX Human: cytostatic; gene therapy; colon cancer; prostate cancer;  
 XX breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX NC020102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000MO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

XX (CHIR) CHIRON CORP.

XX (HSE-) HSE INC.

XX Williams LT, Escobedo J, Imila MN, Garcia PD, Klinger J, Kassam A;  
 XX Reinhard C, Randoz P, Katozky G, Fox D, Lamson G, Drmanac R;  
 XX Crenklov R, Drmanac S, Dickson J, Leishowitz D;  
 XX Klea D, Garcia V, Jones LW, Strachan-Crain B;

XX WPI: 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a  
 PT mammalian cell and detecting cancer, particularly of the colon or  
 PT prostate, comprises 3351 human polynucleotide sequences -  
 PS Claim 9: Page 807; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to screen for detection of mRNA and to produce additional copies  
 CC of the polynucleotide and for detection of the polynucleotide  
 CC or antisense oligonucleotides can be generated from the library  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and  
 CC treat the cancerous state. The polynucleotides and polypeptides can  
 CC be used in pharmaceutical compositions to  
 CC dysplasia and hyperplasia.

XX Sequence 371 BP; 107 A; 79 C; 98 G; 87 T; 0 other;

Query Match 0.8%; Score 37; DB 22; Length 371;  
 DB Local Similarity 100.0%; Pred. No. 6, 3e-05;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 885 gtaggcacacatagtcacgtcgtatagatgag 921

Db 187 gtgagccacaatacgtcactgcgtatagatgag 223  
|||||

## RESULT 8

ID AAF66074 standard: cDNA: 401 BP.

AC AAF66074:

BT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1830.

KW Human: cytostatic; gene therapy; colon cancer; prostate cancer; breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PM WO200102568-A2.

PI 11-JAN-2001.

PF 30-JUN-2000: 2000MO-US18374.

PR 02-JUL-1999: 99US-0142310.

PR 02-JUL-1999: 99US-0142311.

PA (CHIR) CHIRON CORP.

PA (HSE-) HSEB INC.

PI Williams LT, Escobedo J, Innis NA, Garcia PD, Klinger J, Kassam A; Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crtjenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D; Kita D, Garcia V, Jones LW, Strache-Crain B;

PI WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a mammalian cell and detecting cancer, particularly of the colon or prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9: Page 805: 1046pp: English.

XX The present sequence is one of 3351 sequences in a library of human polynucleotides. The library is used to detect differentially expressed genes in a cell with a cancerous state of a mammalian cell and can

CC detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping of

CC the polynucleotides. The probes can be used for chromosome mapping of the polynucleotide and for detection of transcription levels. Ribozymes

CC or antisense oligonucleotides can be generated. The polynucleotides and their gene products are used as genetic or biochemical markers (e.g. in

CC blood or tissues) that will detect the presence of a cancerous state and prevent the cancer from spreading. The polynucleotides can be used in

CC preventive interventions. The polynucleotides, polypeptides and antibodies against them can be used in pharmaceutical compositions to

CC treat the cancers and proliferative disorders such as neoplasia, dysplasia and hyperplasia.

CC Sequence 401 BP: 114 A: 85 C: 107 G: 93 T: 2 other:

SO

Query Match 0.83: Score 37: DB 22: Length 401:

Best Local Similarity 100.0%: Pred. No. 6.2e-05:

Matches 37: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 885 gtgagccacaatacgtcactgcgtatagatgag 921  
|||||

DB 187 gtgagccacaatacgtcactgcgtatagatgag 223

RESULT 9

AAA45014 standard: cDNA: 571 BP.

XX AAA45014:

DE 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:1589.

XX Human: mouse; chicken; rat; secreted expressed sequence tag; sEST;

XX expressed sequence tag; EST; probe; chemotactic; proliferative;

XX immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;

XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antiangi;

XX antiviral; antidiabetic; antistomatitis; vulnary; antiparkinson;

XX anticancer; osteoprotective; neuroprotective; neurotropic; antiparkinson;

XX cerebroprotective; anticonvulsant; antidepressant; antiparkinson;

XX vaccine; autoimmune disorder; antiparkinson; antiparkinson;

XX insulin; cell deficiency; burn; osteoporosis; osteoarthritis;

XX central nervous system disorder; Alzheimer's disease; stroke;

XX Parkinson's disease; Huntington's disease; coagulation disorder;

XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

XX tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999: 99MO-US24206.

XX 15-OCT-1998: 98US-0104436.

XX (GENY) GENETICS INSTR INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX Meberg D, Treacy M, Bowman MR;

XX WPI: 2000-317938/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted

XX expressed sequence tags (sESTs), useful for treating various disorders

XX such as autoimmune, infectious, and central nervous system disorders -

XX Claim 1: Page 616: 803pp: English.

XX AAA45014 to AAA45925 represent specifically claimed secreted expressed

XX sequence tags (sESTs), isolated from human, mouse, chicken and rat

XX tissue sources. The sESTs can have a range of other functions including

XX the tissues they were isolated from. The sESTs can be used in a variety of

XX applications, including: haematopoietic; chemokine; analgesic; haemostatic;

XX chemotactic; antidiabetic; antistomatitis; vulnary; antiparkinson;

XX antiparkinson; osteoprotective; neuroprotective; neurotropic; antiparkinson;

XX anticonvulsant; antidepressant; antiparkinson; antiparkinson;

XX identification and isolation of the sESTs can be used as probes for gene

XX expression and for determining biological activity and raising

XX antibodies. They may be useful for treatment of autoimmune disorders

XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,

XX osteoporosis, osteoarthritis, central nervous system disorders

XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation

XX disorders (hemophilia, thrombosis), inflammation, depression and

XX diseases (tumors, osteoarthritis, Crohn's disease, Crohn's

XX disease), and other disorders. The sESTs represent linker variants which are given

XX in the exemplification of the present invention.

XX Sequence 571 BP: 246 A: 80 C: 125 G: 120 T: 0 other;





AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95883 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 1628 BP: 707 A: 236 C: 348 G: 337 T: 0 other:

Query Match 0.7%; Score 35; DB 22; Length 1628;  
 Best Local Similarity 100.0%; Pred. No. 0.00032;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3660 gaagctttagaagaagctgcatcaaaagc 3694  
 |||||||||||||||||||||||||||||||  
 1585 gaagctttagaagaagctgcatcaaaagc 1619

## RESULT 12

AAAT7915  
 ID AAAT7915 standard; cDNA: 421 BP.

AAAT7915;

14-NOV-2000 (first entry)

cDNA encoding human colon tumour protein CT607, SEQ ID NO:195.

Human colon tumour polypeptide; tumour antigen; cancer; vaccine; immunotherapy; diagnosis; progression; ss.

Homo sapiens.

MO200037643-A2.

29-JUN-2000.

23-DEC-1999; 99MO-US30909.

23-DEC-1998; 98US-0221298.

02-JUL-1999; 99US-0347496.

22-SEP-1999; 99US-0401064.

19-NOV-1999; 99US-0444242.

02-DEC-1999; 99US-0454150.

(CORI-) CORIXA CORP.

Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;

Wang T, Ingliu J;

WPI: 2000-442671/38.

P-PSDB: AAB11903.

Claim 29: Page 147; 22pp; English.

Sequences AAAT7722-A78199 represent 478 cDNAs encoding proteins or portions of proteins which are associated with human colon tumours. The invention also specifically discloses 8 human colon tumour proteins (AAB11897-111904). The nucleic acids, the polypeptides they encode, and antigen presenting cells (APCs), preferably dendritic cells, expressing such polypeptides may be used in vaccine compositions for the development of T-cell specific for the polypeptide expressed by the APC are used to remove tumour cells from biological samples, especially blood or fractions thereof. The sample or the isolated T-cells specific for the polypeptide can then be used to inhibit cancer development. CD4+ and/or CD8+ T-cells from a patient may be incubated with a polypeptide, nucleic acid of the invention, or an APC expressing such a polypeptide, to cause the proliferation of specific T-cells. The inhibit cancer cloned and then administered back to the patient to inhibit cancer

development. Nucleic acids encoding the polypeptides and antibodies against the polypeptides may be used to determine the expression level of a tumour protein of the invention, and therefore to determine whether cancer cells are present. Such diagnostic methods may also be used to monitor the progression of a cancer by repeating the processes at time intervals, and comparing the current result to previous results. The present sequence represents a cDNA encoding a human colon tumour polypeptide.

Sequence 421 BP: 109 A: 101 C: 125 G: 86 T: 0 other:

Query Match 0.7%; Score 32; DB 21; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

617 atacaagagcttcatatcaacacgtgacac 648  
 |||||||||||||||||||||||||||||||  
 346 atacaagagcttcatatcaacacgtgacac 377

## RESULT 13

AAI28653  
 ID AAI28653 standard; cDNA: 421 BP.

AAI28653;

12-OCT-2001 (first entry)

Colon tumour related determined cDNA sequence for CT607.

Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic; gene therapy; vaccine; colonic cancer; ss.

Homo sapiens.

MO200149716-A2.

12-JUL-2001.

29-DEC-2000; 2000MO-US35596.

30-DEC-1999; 99US-0476286.

15-FEB-2000; 2000US-0480321.

06-MAR-2000; 2000US-0504629.

19-MAY-2000; 2000US-0519444.

29-JUN-2000; 2000US-0575251.

28-AUG-2000; 2000US-0609448.

(CORI-) CORIXA CORP.

Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;

King GE, Wang T, Jiang Y;

WPI: 2001-441847/47.

Claim 25: Page 187; 47pp; English.

The present invention describes colon tumour associated proteins (I) and the polynucleotides (II) that encode them. (I) have cytoskeletal activity. (I) and (II) can be used in the prevention, diagnosis and treatment of diseases (I) may be used with inappropriate colon tumour associated protein (TCAP) associated with inappropriate colon tumour associated protein (TCAP) expression, such as colonic cancer. For example, (I) expression may be used to treat disorders associated with inappropriate expression by the colon tumour associated protein. In a patient's genome that affect the activity of TCAPs by expressing inactive proteins or to supplement the patients own production of them. Additionally, (II) may be used to produce the TCAP proteins, by inserting the nucleic acids into a host



CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the 3'-end of the poly(nucleotide) sequence which comprises one of  
 CC the 5602 oligonucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to a  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the combination of  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from antisense therapy and  
 CC the specification. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

S0 Sequence 848 BP: 373 A: 116 C: 164 G: 191 T: 4 other:

Query Match 0.7%; Score 32; DB 22; Length 848;  
 Best Local Similarity 100.0%; Pred. No. 0.005;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2660 aataatcagaaagattggaattagaa 2691  
 Db 385 aataatcagaaagattggaattagaa 416

RESULT 16  
 AAH33568  
 ID AAH33568 standard; cDNA: 653 BP.  
 AC AAH33568;  
 XX  
 XX 03-SEP-2001 (first entry)  
 DT  
 XX  
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:624.  
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma; ss.  
 KW Homo sapiens.  
 MO000122920-A2.  
 PD 05-APR-2001.  
 PF 28-SEP-2000; 2000MO-US26524.  
 PR 29-SEP-1999; 690S-0157137.  
 PR 03-NOV-1999; 990S-0167280.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
 DR MPI: 2001-235357/24.  
 DR P-Psdb: AAG74137.  
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 PS Claim 1: Page 2661: 9803pp; English.

AAH12943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer-associated proteins,  
 CC the proteins are collectively known as colon cancer-associated proteins,  
 CC cancer antigens have cytoskeletal activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders or  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing P.  
 CC Additionally, N may be used to treat the patient's cancer-associated P.  
 CC By expressing the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAH77789 represent sequences used in the exemplification of the  
 CC present invention.

S0 Sequence 653 BP: 216 A: 84 C: 103 G: 243 T: 7 other:

Query Match 0.6%; Score 29; DB 22; Length 653;  
 Best Local Similarity 100.0%; Pred. No. 0.074;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1517 aagaagatagctctcttaaacagatca 1545  
 Db 571 aagaagatagctctcttaaacagatca 599

RESULT 17  
 AA216317  
 ID AA216317 standard; cDNA: 764 BP.  
 AC AA216317;  
 XX  
 XX 12-OCT-1999 (first entry)  
 DT  
 XX  
 DE Human gene expression product cDNA sequence SEQ ID NO:3787.  
 KW Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 KW Homo sapiens.  
 MO9938972-A2.  
 PD 05-AUG-1999.  
 PF 28-JAN-1999; 99MO-US01619.  
 PR 03-APR-1998; 980S-0080666.  
 PR 28-JAN-1998; 980S-0075934.  
 PR 31-MAR-1998; 980S-0080114.  
 PR 03-APR-1998; 980S-0080515.  
 XX (CHIR ) CHIRON CORP.  
 PA (HYSR ) HYSRD INC.  
 PI Crivenjakov R, Dickson M, Dmanac R, Dmanac S,  
 PI Sredobac J, Kasean A, Kennedy J, Glase K, Janis M,  
 PI Lanson G, Ieshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Steche-Crain B, Sudduth-Rlinger J, Williams LT;  
 DR MPI: 1999-494092/41.  
 XX NOVEL human genes and their expression products which are  
 PT differentially expressed in different cell types  
 PS Claim 1: Page 1804: 2479pp; English.

The present invention describes a library of human polynucleotides

CC comprising the sequences given in AA12532 to AA17779. Also described is  
 CC a method of detecting differentiallly expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the polynucleotide sequences given in AA12532 to AA17779. The  
 CC polynucleotides can be used for a variety of purposes, such as  
 CC mapping, tissue typing or profiling, forensic identification, diagnosis  
 CC detection of polymorphisms, polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC encoded genes and to detect differences in expression levels between  
 CC two cells (e.g. to identify normal or diseased tissue in a human,  
 CC cancer). The polynucleotides of the invention are also useful in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.

SO Sequence 764 BP; 194 A; 181 C; 177 G; 198 T; 14 other:

Query Match 0.5%; Score 23; DB 20; Length 764;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4708 tccatcaaaaaaaaaaaaaa 4730  
 Db 741 tccatcaaaaaaaaaaaaaa 763

## RESULT 18

AA059530  
 ID AAC59530 standard; cDNA; 932 BP.

XX AAC59530;

DT 15-FEB-2001 (first entry)

DE Human secreted protein cDNA sequence #24.

KW Cytostatic; immunosuppressive; nocrotropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; anticancer;  
 KW antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW neurotrophic; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; ss.

OS Homo sapiens.

PN WO20005352-A2.

XX 21-SEP-2000.

PF 09-MAR-2000; 2000MO-US06044.

PR 12-MAR-1999; 99US-0124099.

PR 03-DEC-1999; 99US-0168661.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI: 2000-602124/57.

XX P-FSDB: AAB34239.

PT Novel human secreted proteins useful for diagnosis, prevention and  
 PT treatment of disorders including neurological, cell proliferative,  
 PT cardiovascular, autoimmune and inflammatory disorders and microbial  
 PT infections -  
 XX  
 XX  
 XX  
 PA Claim 1; Page 33; 383pp; English.

XX The invention relates to the isolation of genes AAC59507-C59556 encoding  
 XX 50 genes secreted proteins AAB34218-B34264. The genes can be used to  
 XX generate polypeptides and antibodies. The genes can be used to  
 XX immunoglobulin G Fc portion (AAC59548) the gene encoding the antibody of  
 XX the fusion protein as compared to the human protein. The genes and  
 XX proteins are useful for preventing, ameliorating or treating medical  
 XX conditions, e.g. by protein or gene therapy. The genes are isolated  
 XX from a range of human tissues disclosed in the specification. The  
 XX nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 XX and ovarian cancer, (b) infectious diseases, (c) viral infections,  
 XX marrow, breast, gastrointestinal, (d) endocrine disorders, (e)  
 XX hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 XX colitis; (f) cardiovascular disorders such as myocardial ischaemias; (g)  
 XX wound healing; (h) neurological diseases e.g. cerebral anoxia and  
 XX epilepsy; and (i) infectious diseases such as viral, bacterial, fungal  
 XX and parasitic infections.

SO Sequence 932 BP; 256 A; 233 C; 150 G; 291 T; 2 other:

Query Match 0.5%; Score 23; DB 21; Length 932;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4708 tccatcaaaaaaaaaaaaaa 4730  
 Db 885 tccatcaaaaaaaaaaaaaa 907

## RESULT 19

AA077460  
 ID AAC77460 standard; cDNA; 1350 BP.

XX AAC77460;

DT 08-FEB-2001 (first entry)

DE Human ORF3015 polynucleotide sequence SEQ ID NO:6029.

KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;  
 KW viral; anti-inflammatory; antiparasitic; neuroprotective; anticancer;  
 KW anticonvulsant; osteopathic; coagulant; vasculoprotective;  
 KW immunostimulant; thrombolytic; immunosuppressive; antinflammatory;  
 KW hypotensive; dermatological; antineurotic; antitumor;  
 KW antiviral; antibacterial; antifungal; antiparasitic; antitumor;  
 KW antineurotic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cardiovascular disease; diabetes mellitus; erythematous; infection;  
 KW cholestasis; autoimmune disease; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; nocturnal hemoglobinuria; asthma;  
 KW allergy; aplastic anemia; nocturnal hemoglobinuria; infection;  
 KW thrombosis; contraceptive; ss.

OS Homo sapiens.

PN WO200058473-A2.

XX 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX	(HUNA-) HUMAN GENE SCI INC.
PA	Carter KC, Ebner RC, Endress CH, Feng P, Janat F,
P1	Kyam H, Laflaur DM, Leung R, Ni Y, Olsen HS, Rosen CA;
P1	Ruben SM, Shi Y, Soppet DR, Wei Y,
DR	MPI: 1999-337740/28.
DR	P-PDB: AAU27569.
XX	New human secreted proteins and coding sequences useful for treating disorders of the immune system and hyperproliferative disorders
PT	disorders of the immune system and hyperproliferative disorders
XX	Claim 1: page 283; 507pp: English.
XX	
CC	This sequence represents a nucleic acid molecule which encodes a secreted human protein. The gene number is given in the descriptor line. CC
CC	The gene can be used to generate fusion proteins by linking to the gene
CC	to a human immunoglobulin Fc portion (e.g. AKA84924). For the
CC	stability of the fused protein, see examples and their fragments (nucleic
CC	acid sequences: AAK84933-x85057; amino acid sequences AAU27567-y27933)
CC	which are useful for preventing, treating or ameliorating medical
CC	conditions e.g. by protein or gene therapy. Also, pathological
CC	conditions can be diagnosed by determining the amount of the new
CC	polypeptides in a sample or by determining the presence of mutations in
CC	the new polynucleotides. Specific uses are described below in
CC	polymeric materials, based on which tissues they are most highly expressed in
XX	(see AKA84933 for described uses).
SQ	Sequence 1589 BP; 359 A; 379 C; 361 G; 478 T; 12 other:
Query Match	0.5%; Score 23; DB 20; Length 1589;
Best Local Similarity	100.0%; Pred.No.14;
Matches 23: Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	4708 tccatcaaaataaaaaaaaaaaaaa 4700
dbb	1561 tccatcaaaataaaaaaaaaaaaaa 1583
RESULT 21	
AAAB6810	AAA68010 standard: DNA: 1866 BP.
ID	AAA68010;
AC	
XX	24-OCT-2000 (first entry)
XX	Eucalyptus grandis F5H nucleotide sequence SEQ ID NO:103.
XX	plant; lignin biosynthetic pathway; Eucalyptus grandis;
XX	Pinus radiata; Monterey pine; ds.
OS	Eucalyptus grandis.
PN	MO200022099-A1.
PD	20-APR-2000.



PX	Claim 1 SEQ ID NO 269: 1007bp: English.
XX	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	nucleotide sequences thereof (AA157798-AA161369), particularly those
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral neuropathies, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's chorea, amyotrophic
CC	lateral sclerosis, and Shydraseg syndrome. Other uses include the
CC	treatment of immunological disorders, such as immune system suppression,
CC	C active/inhibit activity, chemoclastic/chemokinetic activity, hematologic
CC	and thrombotolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukemias and
CC	C.N.S. disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	Sequence 3160 BP; 941 A: 648 C: 665 G: 906 T: 0 other:
SQ	
DY	Query Match            0.5%; Score 23; DB 22; Length 3160;
	Best Local Similarity   100.0%; Pred. No. 13;
	Matches     23;       Conservative   0; Mismatches   0; Indels   0; Gaps   0
OY	4708 tccctcgtttttttttttttttttttaaataaaat 4730
	Db      3053 tccctcgtttttttttttttttttttaaaaaa 3075
RESULT 25	
AAH46855	
XX	AAH46855 standard; CDNA: 4302 BP.
AC	
XX	AAH46855:
DT	25-SEP-2001 (first entry)
DE	
DD	Human serine/threonine phosphatase encoding cDNA (clone ID HUBCF83).
KM	Serine/threonine phosphatases: neurotropic; neuroprotective; cytoskeletal;
KM	immunomodulatory; immunosuppressive; anti-inflammatory; antibacterial; ss;
KM	anti-HIV; antiparkinsonian; antisticking; antitumor; antifibrotic;
KW	anthelmintic; vincriste; hepatotropic; cerebroprotection; vulnerary;
KW	antiinflammatory; nephrotoxic; gene therapy; vaccine.
XX	
OS	Homo sapiens.
FN	
FM	
FT	Key                  Location/Organifiers
PT	CDS                  166..1161
XX	/**tag= a
PN	WO200155388-A1.
PD	
XX	02-AUG-2001.
XX	17-JAN-2001; 2001NO-USO1395.
PR	31-JAN-2000; 2000US-O17906S.
PR	04-FEB-2000; 2000US-O180658-
PR	02-MAR-2000; 2000US-O186350.
PA	(HUMA-) HUMAN GENOME SCT INC.
XX	
XI	Rosen CA, Barash SC, Ruben SM;
XX	WPI: 2001-476208/51.
DR	P-PSTB: NAB85475.
XX	
PT	Novel proteins of serine/threonine phosphatase family, useful for diagnosing, treating, preventing and/or prognosing distinct the proteins, including cancer, immune response and neuronal disorders

PT Nucleic acid encoding human aminopeptidase P  
 XX  
 XX Claim 13: Page 165-192: 201pp: English.

XX The invention provides serine/threonine phosphatase family polypeptides  
 XX and polynucleotides encoding them. The polypeptides can be expressed by  
 XX genetic engineering methodology. The polypeptides, polynucleotides and  
 XX antibodies are used for diagnosis, prognosis, prevention and treatment  
 XX of neurodegenerative disorders. Infectious diseases, autoimmune  
 XX diseases, allergic reactions, cardiovascular disorders, cerebrovascular  
 XX disorders, renal disorders, reproductive disorders (see A468646 for a  
 XX detailed description of the diseases that can be treated). They are also  
 XX used in the treatment of pathogenesis, activator of T-cells, to induce  
 XX of B-cell responses, to enhance humoral immunity, to increase serum immunoglobulin  
 XX concentrations, to accelerate recovery of aged populations, to induce  
 XX to boost immunoresponsiveness in aged populations, to induce tumour  
 XX regression, to accelerate recovery of aged populations, to induce tumour  
 XX regression, to prevent hair loss, to stimulate growth and  
 XX differentiation of hematopoietic cells and bone marrow cells, for  
 XX supporting cell culture, to modulate mammalian  
 XX characteristics such as body weight, taste, to modulate mammalian  
 XX skin, to modulate mammalian metabolism, to change color, hair color and  
 XX physical state, and as food additive or preservative. The present  
 XX sequence represents a human serine/threonine phosphatase encoding cDNA.  
 XX  
 XX Sequence 4302 BP: 1259 A: 895 C: 910 G: 1238 T: 0 other;

Query Match 0.5%: Score 23: DB 22: Length 4302:  
 Best Local Similarity 100.0%: Pred. No. 13:

Matches 23: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 4708 tccatcaaaaaaaaaaaaaa 4730  
 DB 4279 tccatcaaaaaaaaaaaaaa 4301

RESULT 26  
 AAX23520/C  
 ID AAX23520 standard: DNA: 45546 BP.

AC AAX23520:

DT 23-JUN-1999 (first entry)

DE Human kidney aminopeptidase P genomic DNA fragment 4.

XX Aminopeptidase: human; Amp: gene therapy: treatment: Amp-deficiency;

XX Prenatal diagnosis: angiodysplasia; antihypertensive agent; atherosclerosis;

XX KW proteolysis: waste degradation; additive; immunohistochemistry; ss.

OS Homo sapiens.

PA MO9911799-A2.

XX 11-MAR-1999.

PF 02-SEP-1998: 98NO-US18426.

PR 02-SEP-1997: 97US-0057854.

XX (MED-1) MEDICAL COLLEGE GEORGIA RES INST.

PA Ryan JW, Sprinkle TC, Venema RC:

XX WPI: 1999-205193/17.

PT Nucleic acid encoding human aminopeptidase P  
 XX  
 XX Claim 13: Page 165-192: 201pp: English.

XX This invention describes the isolation of a novel human aminopeptidase P  
 XX (Amp). This protein is used to produce recombinant Amp and can be used  
 XX for gene therapy for treating Amp-deficient patients with hypodysplasia  
 XX and heterozygous Amp deficiency, including prenatal diagnosis (patents and  
 XX defective in Amp are at risk of developing angiodysplasia if treated with  
 XX angiotensin-converting enzyme inhibitor). The product of the invention is  
 XX a case of excessive Amp expression. The product of the invention is  
 XX used to generate transgenic animals, Amp inhibitory sequences and to  
 XX used to detect mutations. Amp inhibitory sequences are useful as  
 XX antihypertensive agents and to prevent or treat arterial  
 XX or atherosclerosis. The structure of Amp is used to design synthetic  
 XX substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal  
 XX imido bonds, can be used to degrade industrial protein feeds to free  
 XX amino acids, to degrade proteinaceous wastes, as additives in enzyme  
 XX bioreactors used to treat malabsorption syndrome and for studying its  
 XX methods to study Amp deficiency.

XX Sequence 45546 BP: 12027 A: 11359 C: 11380 G: 10780 T: 0 other;

Query Match 0.5%: Score 23: DB 20: Length 45546:  
 Best Local Similarity 100.0%: Pred. No. 10:

Matches 23: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 4708 tccatcaaaaaaaaaaaaaa 4730  
 DB 26452 tccatcaaaaaaaaaaaaaa 26430

RESULT 27  
 AAV83939  
 ID AAV83939 standard: DNA: 80595 BP.

AC AAV83939:

DT 03-MAR-1999 (first entry)

DE HC-contig derived from normal human chromosome 10q25.2 region.

XX Yeast artificial chromosome: YAC: probe: eukaryotic chromosome;

XX KW neocentromere; replication: extra-chromosomal element; cell division; artificial chromosome: gene therapy: mar1del(10);

XX human artificial chromosome; transgenic: chromosome 10: 10q25.2; ss.

XX Homo sapiens.

PA MO9851790-A1.

XX 19-NOV-1998.

PF 13-MAY-1998: 98NO-AU00352.

XX 26-AUG-1997: 97AU-0008791.

PR 13-MAY-1997: 97AU-0006784.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

PA Cancilla MR, Choo K, Du Sert D:

XX WPI: 1999-009773/01.

XX New isolated nucleic acid comprising neocentromere sequences from  
 XX eukaryotic chromosomes used to produce replicable, segregating  
 XX artificial chromosomes that can carry large amounts of DNA for gene  
 XX therapy



PS Claim 8: Fig 6: 540bp: English.

CC The present sequence represents the HC-ccnt1g derived from normal human  
 CC chromosome 10, located on the short arm of the chromosome and can be naturally mutated to  
 CC mar10d(10) marker is mitotically stable and contains a functional  
 CC neocentromere at a location regarded as non-centromeric. This  
 CC neocentromere maps to q25.2 on chromosome 10. The specification describes  
 CC nucleic acid sequences derived from a eukaryotic chromosome, including a  
 CC neocentromere or its functional derivative or hybrid, that are able, in  
 CC a compatible cell, of replicating, acting as extra-chromosomal, used to  
 CC control the expression of a gene, or as a gene therapy comprising a  
 CC constructable segregating nucleic acid that confers a specific phenotype  
 CC on cells. Human artificial chromosomes can propagate in human cells and  
 CC carry large amounts of DNA (e.g. therapeutic genes), and, being  
 CC extra-chromosomal, they are not mutagenic. The artificial chromosomes  
 CC are also useful for generation of transgenic plants and animals. In  
 CC production of proteins and to make diagnostic reagents, e.g. for  
 CC expression of cytokines, receptors and of constructs may also be  
 CC used for functional and structural analysis of chromosomes.

XX Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 other;

SO Query Match 0.5%; Score 23; DB: 20; Length 80595;  
 Best Local Similarity 100.0%; Pred. No. 5.7; 0; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0; Gaps 0;

QY 4708 tccatcaaaaaaaaaaaaaa 4730  
 DB 25198 tccatcaaaaaaaaaaaaaa 25220

RESULT 28  
 AAT69845/C  
 ID AAT69845 standard; DNA: 29 BP.

AC AAT69845;  
 DT 05-MAR-1998 (first entry)  
 XX Address probe #2 for mutant ornithine transcarbamylase gene.  
 DE  
 XX Open circle probe: rolling circle replication primer; mutation detection;  
 KM amplification target circle; rolling circle replication amplification;  
 KM multiple target detection; circular DNA; ss.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX MO9719193-A2.  
 PN 29-MAY-1997.  
 XX 21-NOV-1996; 96KO-US18612.  
 XX 01-MAY-1996; 96US-0016677.  
 PR 21-MAY-1995; 95US-0563912.  
 XX (UYVA ) UNIV YALE.  
 PA Caplan M, Lizardi PM;  
 XX WPI: 1997-298125/27.  
 XX Unimolecular segment amplification method for amplifying nucleic  
 PT acids - uses rolling circle mechanism of amplification, useful for  
 PT disease-associated genetic mutation(s)  
 XX Example 4: Page 79; 149pp: English.

CC AAT69845 and AAT69846 are address probes for mutant and wild type  
 CC ornithine transcarbamylase, respectively. These sequences were used to  
 CC test the method of the invention. The method of the invention is for  
 CC amplifying a nucleic acid molecule. It comprises mixing at least one  
 CC rolling circle replication primer (RCRP) with at least one amplification  
 CC target circle (ATC) which comprises a single stranded circular DNA  
 CC containing a region complementary to the RCRP, such as with the  
 CC hybridises to the ATC and undergoes a tandem sequence DNA is formed. The  
 CC polymerase designated unimolecular segment amplification, is used for  
 CC detecting specific sequences with high speed and sensitivity. It is  
 CC especially useful for detecting mutations in genes where numerous  
 CC distinct mutations are associated with a disease or where mutations in  
 CC multiple genes are involved. The method is isothermal and provides a  
 CC simple, consistent and quantitative detection of multiple target  
 CC molecules as the amplification occurs in a single reaction. The  
 CC continuous reaction: the amplification is highly quantitative as the  
 CC amplification step is linear and is catalysed by a highly processive  
 CC enzyme.

XX Sequence 29 BP; 5 A; 0 C; 6 G; 18 T; 0 other;

SO Query Match 0.5%; Score 22; DB: 16; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 5; 0; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 0; Gaps 0;

QY 4708 tccatcaaaaaaaaaaaaaa 4729  
 DB 22 tccatcaaaaaaaaaaaaaa 1

RESULT 29  
 AAT61892/C  
 ID AAT61892 standard; DNA: 255 BP.

AC AAT61892;  
 DT 21-SEP-2001 (first entry)  
 XX Rat differential transcription-associated cDNA SEQ ID 401.  
 DE  
 XX Differential transcription: human; rat; tumour cell; cytosolic;  
 KM Ras modulator; Class II tumour suppressor gene; gene therapy; ss.  
 OS Rattus sp.  
 XX MO200157058-A2.  
 PN 09-AUG-2001.  
 XX 31-JAN-2001; 2001WO-EP01003.  
 XX 31-JAN-2001; 2000DE-1004102.  
 XX (METHA-) METAGEN GES GENOMFORSCHUNG MBH.  
 PA Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;  
 PI Grips M, Hellriegel M, Schmitz A, Sers C;  
 XX WPI: 2001-483415/52.  
 XX Nucleic acid differentially expressed between tumor and normal cells,  
 PT useful for diagnosis or therapy of tumors and for screening active  
 PT agents -  
 XX Claim 6: Page 490; 579pp: German.

CC This invention describes a nucleic acid (1) with differential expression  
 CC between tumour cells and which has cytostatic activity. (1)  
 CC between modulators of Ras activity by inducing expression of tumour  
 CC suppressor genes. (1), and polypeptides encoded by them, are useful as  
 CC targets for diagnosis or therapy and in screening to determine the



CC (1) can be used for monitoring the progression of cancer in a patient

[illegible]



```

XX 08-DEC-1999; 9905-0169681.
PR 21-DEC-1999; 9905-0171350.
PR 14-MAR-2000; 2000US-0193735.
PR 07-MAR-2000; 2000US-0193735.
PR 07-MAR-2000; 2000US-0210660.
PR 21-JUL-2000; 2000US-0220114.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI: 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
XX and for assessing and detecting compounds for treating the cancer -
XX
XX Claim 1: Page 177-178; 1051pp; English.
XX
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
XX cervical cancer with cytostatic activity. The nucleic acids encoded
XX polypeptides are useful: (a) as a diagnostic condition to monitor the
XX progression of cervical cancer; (b) as a prognostic condition in a patient;
XX (c) to select and/or assess the efficacy of a compound or therapy for
XX inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy.
XX
XX Sequence 367 BP; 123 A; 64 C; 78 G; 102 T; 0 other:
XX
XX Query Match 0.5%: Score 22; DB 22; Length 367;
XX Best Local Similarity 100.0%; Pred. No. 40;
XX Matches 22; Conservative 0; Mismatches 0; Gaps 0;
XX
XX QY 4709 ccctcaaaaaaaaaaaaaa 4730
XX |||||||||||||||||||
XX DB 335 ccatcaaaaaaaaaaaaaa 356
XX
XX RESULT 37
XX AAH68305
XX ID AAH68305 standard; cDNA; 407 BP.
XX
XX AAH68305;
XX
XX 12-FEB-1999 (first entry)
XX
XX EST clone D0410.
XX
XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
XX chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytics;
XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
XX
XX M09845437-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98MO-US06956.
XX
XX 10-APR-1997; 97US-0837312.
XX
XX (GENE) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Werberg D;
XX PI Racle LA, Spaulding V, Treacy M;
XX WPI: 1999-070078/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from

```

```

PR e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pleurotary, retina and colon cDNA libraries
XX
XX Claim 1: Page 351; 641pp; English.
XX
XX The present sequence represents an expressed sequence tag (EST), and is
XX a polynucleotide of the invention. The polynucleotides of human tissue
XX are all secreted EST sequences isolated from cells which are predicted to
XX sources. The EST sequences are useful: (a) as a diagnostic condition to
XX monitor the progression of cervical cancer; (b) as a prognostic condition
XX in a patient; (c) to select and/or assess the efficacy of a compound or
XX therapy for inhibiting cervical cancer in a patient. The nucleic acids may
XX also be useful for gene therapy.
XX
XX Sequence 407 BP; 177 A; 64 C; 91 G; 75 T; 0 other:
XX
XX Query Match 0.5%: Score 22; DB 20; Length 407;
XX Best Local Similarity 100.0%; Pred. No. 40;
XX Matches 22; Conservative 0; Mismatches 0; Gaps 0;
XX
XX QY 3093 taaagctgagatcacagcct 3114
XX |||||||||||||||||||
XX DB 146 taaagctgagatcacagcct 169
XX
XX RESULT 38
XX AAH94999
XX ID AAH94999 standard; DNA; 489 BP.
XX
XX AAH94999;
XX
XX 23-MAY-2001 (first entry)
XX
XX Human ovarian cancer associated coding sequence SEQ ID NO: 191.
XX
XX Human, ovarian cancer; vaccine; gene therapy; carcinoma; ds.
XX
XX Homo sapiens.
XX
XX M0200118046-A2.
XX
XX 15-MAR-2001.
XX
XX 08-SEP-2000; 2000MO-US24827.
XX
XX 10-SEP-1999; 99US-0394374.
XX
XX 01-MAY-2000; 2000US-0561778.
XX
XX 15-AUG-2000; 2000US-0640173.
XX
XX 07-SEP-2000; 2000US-0656668.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Stolk JA;
XX WPI: 2001-211395/21.
XX
XX Isolated polypeptides associated with ovarian carcinomas, and the
XX nucleic acids that encode them, useful for the prevention, diagnosis and
XX treatment of ovarian cancers -
XX
XX Claim 18; Page 186-187; 189pp; English.
XX
XX The present invention provides a number of coding sequences and proteins,
XX the over-expression of which is associated with ovarian carcinoma/cancer.
XX These can be used in the diagnosis, treatment and prevention of ovarian
XX cancer, optionally by gene therapy or in the form of a vaccine. The

```

CC present sequence is an example of one of these sequences.  
 XX Sequence 489 BP; 186 A; 65 C; 69 G; 167 T; 2 other:  
 SQ

Query Match 0.5%: Score 22; DB 22; Length 489;  
 Best Local Similarity 100.0%; P-Val No. 39;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatcaaaaaaaaaaaaaa 4730  
 Db 463 ccatcaaaaaaaaaaaaaa 484

RESULT 39  
 AA140146/c  
 ID AA140146 standard; DNA; 577 BP.  
 AC AA140146;

XX 17-OCT-2001 (first entry)

DE Probe #832 used to measure gene expression in human placenta sample.  
 KW Probe: microarray; human; placenta; antenatal diagnosis;  
 KM genetic disorder; ss.

XX Homo sapiens.

XX M0200157272-42.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-MAR-2000; 2000US-0208408.

XX 03-AUG-2000; 2000US-0234685.

XX 21-SEP-2000; 2000US-0236355.

XX 27-SEP-2000; 2000US-0236355.

XX 04-OCT-2000; 2000CB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-468897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID NO 8832; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

XX The present sequence is the South probe. The probes are useful for

XX CC producing a microarray for

XX CC expression in samples derived from human

XX CC for antenatal diagnosis of human genetic disorders.

XX Sequence 577 BP; 189 A; 101 C; 106 G; 181 T; 0 other:

Query Match 0.5%: Score 22; DB 22; Length 577;  
 Best Local Similarity 100.0%; P-Val No. 38;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4709 ccatcaaaaaaaaaaaaaa 4730

Db 179 CCNTCAAAAAAAAAAAAAA 158

RESULT 40

MAC53962  
 11 MAC53962 standard; DNA; 584 BP.  
 AC MAC53962;

XX 18-OCT-2000 (first entry)

XX Zea mays DNA fragment SEQ ID NO: 76203.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein production; signal transduction pathway; metabolic;

XX pathway; promoter; termination sequence; corn; ss.

XX Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 23-MAR-1999; 99US-0126784.

XX 01-MAR-1999; 99US-0126785.

XX 06-APR-1999; 99US-0128214.

XX 08-APR-1999; 99US-0128214.

XX 16-APR-1999; 99US-0128445.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 25-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132444.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132487.

XX 14-MAY-1999; 99US-0132487.

XX 14-MAY-1999; 99US-0132487.

XX 18-MAY-1999; 99US-0132487.

XX 19-MAY-1999; 99US-0132487.

XX 20-MAY-1999; 99US-0132487.

XX 21-MAY-1999; 99US-0132487.

XX 25-MAY-1999; 99US-0132487.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 08-JUN-1999; 99US-0137724.

XX 10-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138477.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 17-JUN-1999; 99US-0139452.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139464.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139817.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0142820.  
 PR 01-JUL-1999; 99US-0142821.  
 PR 01-JUL-1999; 99US-0142822.  
 PR 02-JUL-1999; 99US-0142823.  
 PR 06-JUL-1999; 99US-0142824.  
 PR 08-JUL-1999; 99US-0142825.  
 PR 09-JUL-1999; 99US-0142826.  
 PR 12-JUL-1999; 99US-0142827.  
 PR 13-JUL-1999; 99US-0142828.  
 PR 13-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 20-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144332.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 26-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145225.  
 PR 27-JUL-1999; 99US-0145226.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147008.  
 PR 03-AUG-1999; 99US-0147009.  
 PR 04-AUG-1999; 99US-0147010.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147535.  
 PR 10-AUG-1999; 99US-0148110.  
 PR 11-AUG-1999; 99US-0148111.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149428.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149925.

PR 23-AUG-1999; 99US-0149920.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151080.  
 PR 31-AUG-1999; 99US-0151103.  
 PR 31-AUG-1999; 99US-0151130.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 20-SEP-1999; 99US-0154780.  
 PR 22-SEP-1999; 99US-0154782.  
 PR 23-SEP-1999; 99US-0154783.  
 PR 24-SEP-1999; 99US-0154784.  
 PR 28-SEP-1999; 99US-0156456.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157153.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158032.  
 PR 12-OCT-1999; 99US-0158364.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161389.  
 PR 26-OCT-1999; 99US-0161380.  
 PR 26-OCT-1999; 99US-0161380.  
 PR 26-OCT-1999; 99US-0161380.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match Similarity 100.0%; Score 22; DB 21; Length 584;  
 Best Local Similarity 100.0%; Pred No. 38;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4709 cccctcaaaaaaaaaaaaaa 4730  
 Db 540 cccctcaaaaaaaaaaaaaa 561

RESULT 41

AAAG4589 standard; DNA: 589 BP.

AAAG4589.

XX 02-JAN-2001 (first entry)  
 XX  
 DT  
 XX

DE Nucleotide sequence of a human polynucleotide.

XX OLD-35; OLD-64; OLD-137; OLD-139; OLD-142; OLD-145; cancerous phenotype;  
 KW cellular senescence; terminal differentiation; growth suppression;  
 KM aging process; type I interferon; cancer cell; tissue regeneration; ss.  
 XX  
 OS Homo sapiens.

FN M0200046391-A2.

PD 10-AUG-2000.

PK 02-FEB-2000; 2000MO-US02920.

PR 02-FEB-1999; 99US-0243277.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

PA  
 PI Fisher PB, Leszczynska M;  
 XX MPI: 2000-532905/48.

PT Novel isolated nucleic acid encoding an OLD-35 or OLD-64 protein useful  
 PT in the treatment and detection of e.g. cancer and diseases involving  
 PT cellular senescence -

PS Disclosure; Page 97; 115pp; English.

XX The specification describes OLD-35, OLD-64, OLD-137, OLD-139, OLD-142  
 CC OLD-145 proteins. The OLD nucleic acids are useful for reversing  
 CC the cancerous phenotype of a cancer cell, determining if a cell is  
 CC senescent, growth arrested, or differentiating. They are also  
 CC useful for reversing the aging process and for upregulating specific  
 CC RNAs in a cell. The genes may also be used as a diagnostic tool for  
 CC cellular senescence, terminal differentiation and/or growth suppression  
 CC and as a marker to identify drugs or small molecules that will induce  
 CC or inhibit cellular senescence or terminal differentiation and type I  
 CC is useful for reversing the differentiation of specific cells. Old-35  
 CC Fitch 3' UTRs. The OLD proteins are useful for inducing cancerous  
 CC phenotype of a cancer cell and inhibiting the growth of a cancer cell.  
 CC They are also useful for regenerating tissue. The present sequence  
 CC appears in the specification.

XX  
 XX  
 SO Sequence 589 BP; 198 A; 127 C; 120 G; 141 T; 3 other:

Query Match 0.5%; Score 22; DB 21; Length 589;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4709 cctccaaaataaaaaaataa 4730  
 |||||||  
 DB 531 cctccaaaataaaaaaataa 552

RESULT 42  
 AAV34397/C  
 XX AAV34397 standard; DNA; 843 BP.  
 AC AAV34397;  
 XX  
 DT 02-MAR-1999 (first entry)  
 XX  
 DE Sunflower albumin gene 5' regulatory region.  
 XX  
 KW Upstream region; regulatory region; sunflower; albumin; seed; expression;  
 KM lipid metabolism; delta-6 desaturase; transgenic plant; ds.  
 XX Helianthus annuus.  
 CS Synthetic.  
 XX

PH Key Location/Qualifiers  
 FT misc\_RNA  
 FT 809  
 FT /tag= a  
 FT /note= [transcriptional start site\*]  
 XX  
 PN M09845460-A1.

PD 15-OCT-1998.

PK 09-APR-1998; 98MO-US07178.

PR 09-APR-1997; 97US-0831570.

XX (RHON ) RHONE-POULENC AGRICULTURE.

PA  
 PI Beremand PD, Numberg AN, Thomas TL;  
 XX MPI: 1998-583201/49.

PT New sunflower albumin 5' regulatory region - useful for directing  
 PT altered lipid metabolism in plant seeds

PS Claim 1; Fig 4; 38pp; English.

XX This sequence corresponds to an amplified fragment of the upstream  
 CC region of the sunflower albumin gene 5' regulatory region. The sequence  
 CC spans from -860 to +28 of the albumin gene (Ha5) and is used to direct  
 CC 1997 level seed-specific expression of a heterologous gene, such as  
 CC lipid biosynthetic genes, e.g. delta-6 desaturase (AAV34398), in transgenic  
 CC plants. The 3' BamHI site and the PstI recognition and cleavage site  
 CC (GGATCC) was added during the PCR amplification procedure.

XX  
 XX  
 SO Sequence 843 BP; 261 A; 177 C; 127 G; 278 T; 0 other:

Query Match 0.5%; Score 22; DB 19; Length 843;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4709 cctccaaaataaaaaaataa 4730  
 |||||||  
 DB 306 CCATCAAAAAAAAAAAAAAAAAA 285

RESULT 43  
 AAV76268  
 ID AAV76268 standard; cDNA; 1098 BP.  
 XX  
 AC AAV76268;  
 XX  
 DT 25-JAN-2001 (first entry)  
 XX  
 DE Maize glutathione-S-transferase coding sequence fragment SEQ ID NO: 35.  
 KW Maize; glutathione-S-transferase; GST; plant detoxification;  
 KM herbicide metabolism; ss.  
 XX  
 OS Zea mays.  
 XX  
 PN US6096504-A.  
 XX  
 PD 01-AUG-2000.  
 XX  
 PK 10-FEB-1999; 99US-0248335.  
 PR 05-SEP-1997; 97US-0924759.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI McConlgle B, O'Keefe DP;  
 XX MPI: 2000-523870/47.  
 DR P-PSDB: AAB22117.



XX New nucleic acid fragment encoding maize glutathione-S-transferase  
 PT enzyme for detoxifying xenobiotic compounds in plants and seeds,  
 PT comprises a specific nucleotide sequence -  
 XX  
 XX  
 XX Claim 2: Column 69-72: 62pp: English.  
 CC The present invention concerns the isolation of a number of maize  
 CC glutathione-S-transferase (GST) fragments and their coding sequences. The  
 CC coding sequences were isolated from a number of different clones, leading  
 CC to the identification of a number of different protein sequences. The  
 CC protein is involved in the detoxification of xenobiotic compounds in  
 CC plants and seeds, and has been employed to produce herbicide-resistant  
 CC transgenic plants. The gene and protein can also be used in screening for  
 CC GST inhibitors and the identification of GST substrates, to aid in the  
 CC mapping of plant genomes and as signal sequences to direct molecules to  
 CC the plant vacuole.  
 CC  
 SO Sequence 1098 BP; 368 A; 213 C; 238 G; 279 T; 0 other;  
 Query Match 0.5%; Score 22; DB 21; Length 1098;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatacaaaaaaaaaaaaaa 4730  
 Db 1005 ccatacaaaaaaaaaaaaaa 1026  
 RESULT 44  
 AA159602/C  
 ID AA159602 standard; cDNA: 1233 BP.  
 XX  
 XX AA159602;  
 AC 22-OCT-2001 (first entry)  
 DT Human polynucleotide SEQ ID NO 3591.  
 XX  
 XX Human: noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;  
 KW chemokine; chromolytic; drug screening; arthritis; inflammation;  
 KW leukemia; ss.  
 XX Homo sapiens.  
 XX  
 XX MO200153312-A1.  
 PN  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 26-DEC-2000; 2000MO-US34263.  
 PF  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 29-OCT-2000; 2000US-0693036.  
 PR 19-NOV-2000; 2000US-0727344.  
 XX  
 XX (HNSF-) HYSEQ INC.  
 XX  
 PI Tang Y, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ben F, Wang D;  
 PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao Q, Zhou P, Goodrich R, Dimenac ET;  
 XX WPI: 2001-44235/47.  
 DR P-PSDB: AA04046.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PT  
 XX  
 XX Claim 1: SEQ ID NO 3591; 10078pp: English.  
 PS  
 CC The invention relates to human nucleic acids (AA157796-AA161365) and  
 CC the encoded polypeptides (AA08642-AA042213), with noctropes are useful  
 CC immunosuppressant and cytostatic compounds, a polypeptide of polynucleotide  
 CC in gene therapy, may be used to treat diseases of the peripheral nervous  
 CC system such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilization of the activities such as: immune system suppressive  
 CC activation/inhibition activity, chemotactic activity, haemostatic  
 CC and thrombolytic activity, chemotactic diagnosis and therapy, drug screening,  
 CC C-N-S disorders  
 CC C-N-S disorders  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 CC  
 SO Sequence 1233 BP; 392 A; 226 C; 289 G; 326 T; 0 other;  
 Query Match 0.5%; Score 22; DB 22; Length 1233;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatacaaaaaaaaaaaaaa 4730  
 Db 185 CCATCAAAAAAAAAAAAAAAAAA 165  
 RESULT 45  
 AA241405  
 ID AA241405 standard; cDNA: 1478 BP.  
 XX  
 XX AA241405;  
 AC 19-JAN-2000 (first entry)  
 DT Human normal pancreas tissue derived cDNA 20.  
 XX  
 XX Human: pancreas; cancer; treatment; anticancer; cytostatic;  
 KW gene therapy; EST; expressed sequence tag; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX DE19818598-A1.  
 PN  
 XX 21-OCT-1999.  
 XX  
 PF 19-APR-1998; 98DE-1016598.  
 PR 19-APR-1998; 98DE-1016598.  
 PR  
 XX (META-) METAGEN GBS GENOMESCHUNG MBH.  
 XX  
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 XX WPI: 1999-592039/51.  
 DR  
 PT New nucleic acid sequences expressed in normal pancreatic tissues, and  
 PT derived polypeptides, for treatment of pancreatic cancer and  
 PT identification of therapeutic agents -  
 PT  
 XX Claim 3: Page 59; 92pp: German.  
 XX  
 CC This invention describes novel cDNA sequences (A) that are highly  
 CC expressed in normal human pancreatic tissue and which have anticancer  
 CC and cytostatic activity. (A) are used (1) for recombinant expression of

CC polypeptides (B) and (11) to isolate complete genes. (B) are used (1) to  
 CC identify agents suitable for treatment of pancreatic cancer. (11) are  
 CC directly for creating this form of cancer (including expression from  
 CC gene therapy vectors) and (11) for generation of specific antibodies.  
 CC (A) are identified by assembling ESTs (expressed sequence tags) from a  
 CC tissue of unknown type before comparison of expression patterns. This  
 CC allows a single library of ESTs to be compared with the fact that ESTs  
 CC should reduce the number of false identifications of the gene to be revealed, so  
 CC from different libraries may represent different genes. Of the same  
 CC unknown gene, distorting the estimated frequency of occurrence in a  
 CC particular tissue. AA21386-241423 represent human normal pancreatic  
 CC tissue derived cDNA fragments which encode the protein fragments  
 CC represented in AA93893-Y59920.  
 XX

50 Sequence 1478 BP; 436 A; 251 C; 334 G; 457 T; 0 other;

Query Match

0.5%; Score 22; DB 20; Length 1478;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4709 ccatcaaaaaaaaaaaaaaa 4730

Db 1408 ccatcaaaaaaaaaaaaaaa 1429

Search completed: April 10, 2002, 18:08:41  
 Job time: 7880 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd

Run on: April 10, 2002, 16:06:06 ; Search time 77.85 Seconds  
(without adjustments)

Title:	US-09-750-590-1
Perfect score:	4730
Sequence:	1 caatgtttgagcgcgcaagt.....atcaaaaaaaaaaaaaaa 4730

Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters:	702406
--	--------

Maximum DB seq length: 20000000000

post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:4

- ```

1: /cgn2_6/prodata/2/1na/5A_COMB.seq.*
2: /cgn2_6/prodata/2/1na/5B_COMB.seq.*
3: /cgn2_6/prodata/2/1na/5A_COMB.seq.*
4: /cgn2_6/prodata/2/1na/5B_COMB.seq.*
5: /cgn2_6/prodata/2/1na/PCRTS_COMB.seq.*
6: /cgn2_6/prodata/2/1na/backfiles1.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID                | Description      |
|------------|-------|-------------|--------|----|-------------------|------------------|
| 1          | 23    | 0.5         | 1907   | 3  | US-08-826-611-1   | Sequence 1, App1 |
| 2          | 23    | 0.5         | 8059   | 5  | US-09-078-294-3   | Sequence 3, App1 |
| 3          | 22    | 0.5         | 29     | 2  | US-08-753-912-17  | Sequence 11, App |
| 4          | 22    | 0.5         | 29     | 3  | US-08-754-601-17  | Sequence 17, App |
| 5          | 22    | 0.5         | 29     | 3  | US-08-754-601-17  | Sequence 11, App |
| 6          | 22    | 0.5         | 29     | 4  | US-09-132-552-17  | Sequence 1, App  |
| 7          | 22    | 0.5         | 29     | 4  | US-09-132-552-17  | Sequence 2, App1 |
| 8          | 22    | 0.5         | 8626   | 4  | US-08-926-135-35  | Sequence 3, App  |
| 9          | 22    | 0.5         | 1068   | 3  | US-08-926-135-35  | Sequence 3, App  |
| 10         | 22    | 0.5         | 1784   | 3  | US-08-991-426-1   | Sequence 1, App1 |
| 11         | 22    | 0.5         | 1784   | 3  | US-09-113-470-1   | Sequence 1, App1 |
| 12         | 22    | 0.5         | 4033   | 1  | US-08-116-587C-8  | Sequence 8, App1 |
| 13         | 21    | 0.4         | 232    | 2  | US-08-619-5428-26 | Sequence 26, App |
| 14         | 21    | 0.4         | 522    | 1  | US-08-289-2478-1  | Sequence 1, App1 |
| 15         | 21    | 0.4         | 522    | 5  | PC1-0595-92762-1  | Sequence 9, App1 |
| 16         | 21    | 0.4         | 775    | 4  | US-08-904-731-11  | Sequence 9, App1 |
| 17         | 21    | 0.4         | 775    | 4  | US-08-749-659-9   | Sequence 11, App |
| 18         | 21    | 0.4         | 775    | 4  | US-08-749-659-9   | Sequence 9, App  |
| 19         | 21    | 0.4         | 1055   | 1  | US-08-702-344-18  | Sequence 18, App |
| 20         | 21    | 0.4         | 1055   | 1  | US-08-725-532A-2  | Sequence 2, App  |
| 21         | 21    | 0.4         | 1368   | 3  | US-08-707-359E-1  | Sequence 1, App1 |
| 22         | 21    | 0.4         | 1393   | 3  | US-08-525-597-2   | Sequence 2, App1 |
| 23         | 21    | 0.4         | 1770   | 1  | US-08-032-205-4   | Sequence 3, App1 |
| 24         | 21    | 0.4         | 1770   | 1  | US-08-595-974-3   | Sequence 3, App1 |
| 25         | 21    | 0.4         | 1470   | 1  | US-08-872-302-1   | Sequence 11, App |
| 26         | 21    | 0.4         | 1470   | 1  | US-08-872-302-1   | Sequence 11, App |
| 27         | 21    | 0.4         | 1587   | 3  | US-09-108-020-11  | Sequence 11, App |

|    |    |     |      |   |                   |                   |
|----|----|-----|------|---|-------------------|-------------------|
| 28 | 21 | 0.4 | 1736 | 4 | US-09-360-197-13  | Sequence 13, App1 |
| 29 | 21 | 0.4 | 5459 | 1 | US-07-095-664-4   | Sequence 4, App1  |
| 30 | 21 | 0.4 | 5459 | 1 | US-08-141-887-4   | Sequence 4, App1  |
| 31 | 21 | 0.4 | 5459 | 1 | US-07-059-664-2   | Sequence 2, App1  |
| 32 | 21 | 0.4 | 5659 | 1 | US-08-141-887-2   | Sequence 2, App1  |
| 33 | 20 | 0.4 | 43   | 4 | US-08-306-290-37  | Sequence 37, App1 |
| 34 | 20 | 0.4 | 2603 | 3 | US-08-688-988-19  | Sequence 19, App1 |
| 35 | 20 | 0.4 | 282  | 1 | US-08-550-735-10  | Sequence 10, App1 |
| 36 | 20 | 0.4 | 282  | 4 | US-08-906-480-10  | Sequence 10, App1 |
| 37 | 20 | 0.4 | 371  | 1 | US-08-664-956B-35 | Sequence 23, App1 |
| 38 | 20 | 0.4 | 371  | 2 | US-08-153-775-3   | Sequence 3, App1  |
| 39 | 20 | 0.4 | 532  | 4 | US-08-906-480-1   | Sequence 1, App1  |
| 40 | 20 | 0.4 | 615  | 1 | US-08-539-504-1   | Sequence 1, App1  |
| 41 | 20 | 0.4 | 731  | 1 | US-08-451-605A-2  | Sequence 2, App1  |
| 42 | 20 | 0.4 | 733  | 1 | US-07-512-500-23  | Sequence 23, App1 |
| 43 | 20 | 0.4 | 733  | 1 | US-08-283-509-23  | Sequence 23, App1 |
| 44 | 20 | 0.4 | 733  | 1 | US-08-513-075A-27 | Sequence 27, App1 |
| 45 | 20 | 0.4 |      |   |                   |                   |

## ALIGNMENTS

RESULT 1  
US-08-826-611-1  
; Sequence 1, Application US/08826611

GENERAL INFORMATION:

APPLICANT: Bennett, Alan B.

APPLICANT: Kanayama, Yoshinori

TITLE OF INVENTION: Fructokinase Genes and Their Use In

```

; TITLE OF INVENTION: Me
; NUMBER OF SEQUENCES: 6
;

```

;;  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Crew Life  
;;  
ADDRESS: Two Embroidery Center Fifth Floor  
;;  
CITY: New York, New York

STREET: TWO EMDEN  
CITY: San Francisco  
STATE: California

COUNTRY: USA  
ZIP: 94111-3834

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE:  PatentIn Release #1.0,
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  IIS/08/826.611

```

APPROVAL NUMBER: 007  
FILING DATE: 05-APR-1997  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.

REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-077400U5

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
FAX: (415) 576-0300

TELEFAX: (413) 578-0300  
INFORMATION FOR SEQ ID NO: 1  
SEQUENCE CHARACTERISTICS:

SEQUENCE: 1907 base pairs  
LENGTH: 1907 base pairs  
TYPE: nucleic acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA
; FEATURE:
;

```

```

;
; NAME/KEY: CDS
;
; LOCATION: 271..1314
;
; OTHER INFORMATION: /proc

```

US-08-826-611-1

| Query Match | 0.58; | Score 23; | DB 3; | Length 1907 |
|-------------|-------|-----------|-------|-------------|
|-------------|-------|-----------|-------|-------------|

```
Best Local Similarity 100.0%; Pred. No. 3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

QY 4708 tccatcaaaaaaaaaaaaaa 4730  
 Db 1882 tccatcaaaaaaaaaaaaaa 1904

## RESULT 2

US-09-078-294-3  
 Sequence 3, Application US/09078294  
 Patent No. 6265211  
 GENERAL INFORMATION:  
 APPLICANT: Duoc, Kong-Hong Andy  
 APPLICANT: Duoc, Kong-Hong Andy  
 APPLICANT: Cancelli, Michael R  
 TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
 FILE REFERENCE: Davies Col  
 CURRENT APPLICATION NUMBER: US/09/078,294  
 NUMBER OF SEQ ID NOS: 29  
 SEQ ID NOS: Patentin Ver. 2.0  
 LENGTH: 80595  
 TYPE: DNA  
 ORGANISM: Nucleotide sequence of HC-contig  
 US-09-078-294-3

Query Match 0.5%, Score 23; DB 4; Length 80595;  
 Best Local Similarity 100.0%; Pred. No. 2.3;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4708 tccatcaaaaaaaaaaaaaa 4730  
 Db 25198 tccatcaaaaaaaaaaaaaa 25220

## RESULT 3

US-08-563-912-17/C  
 Sequence 17, Application US/08563912  
 Patent No. 5851033  
 GENERAL INFORMATION:  
 APPLICANT: Lizardi, Paul M.  
 TITLE OF INVENTION: Rolling Circle Replication Reporter Systems  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS: Pabst  
 ADDRESS: 1201 West Peachtree Street  
 STREET: 1201 West Peachtree Street  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30306-3450  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PC/XT/AT/PS/2  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/563,912  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATORNEY/AGENT INFORMATION:  
 NAME: Lizardi, Paul M.  
 REGISTRATION NUMBER: 31,284  
 TELEPHONE: (404)873-8794  
 TELECOMMUNICATION NUMBER: YU115  
 TELEFAX: (404)873-8795  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 29 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single

TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 HYDROLYZABLE: NO  
 ANTI-SENSE: NO  
 US-08-563-912-17

Query Match 0.5%, Score 22; DB 2; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 10;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4708 tccatcaaaaaaaaaaaaaa 4729  
 Db 22 tccatcaaaaaaaaaaaaaa 1

## RESULT 4

US-08-754-681-17/C  
 Sequence 17, Application US/08754681  
 Patent No. 6148195  
 GENERAL INFORMATION:  
 APPLICANT: Lizardi, Paul M. and Caplan, Michael  
 TITLE OF INVENTION: Unimolecular Segment Amplification  
 TITLE OF INVENTION: And Sequencing  
 NUMBER OF SEQUENCES: 28  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: 1201 West Peachtree Street  
 STREET: 1201 West Peachtree Street  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30306-3450  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PC/XT/AT/PS/2  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/754,681  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATORNEY/AGENT INFORMATION:  
 NAME: Lizardi, Paul M.  
 REGISTRATION NUMBER: 31,284  
 TELEPHONE: (404)873-8794  
 TELECOMMUNICATION INFORMATION: YU115C1P2  
 TELEFAX: (404)873-8795  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 29 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-754-681-17

Query Match 0.5%, Score 22; DB 3; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4708 tccatcaaaaaaaaaaaaaa 4729  
 DB 22 tccatcaaaaaaaaaaaaaa 1

RESULT 5  
 US-09-132-552-17/C  
 Sequence 17, Application US/09132552  
 Patent No. 6183960  
 GENERAL INFORMATION:  
 APPLICANT: Lizardi, Paul M.  
 TITLE OF INVENTION: Rolling Circle Replication Reporter Systems  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS: Patent  
 ADDRESS: 2800 One Atlantic Center  
 STREET: 1201 West Peachtree Street  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30306-3450  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/132,552  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/563,912  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pabst, Patrea L.  
 REGISTRATION NUMBER: 31,284  
 REFERENCE/DOCKET NUMBER: YU115  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404)873-8794  
 TELEFAX: (404)873-8794  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 29 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-09-132-552-17

Query Match 0.5%; Score 22; DB 4; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 4708 tccatcaaaaaaaaaaaaaa 4729  
 DB 22 tccatcaaaaaaaaaaaaaa 1

RESULT 6  
 US-09-132-553-17/C  
 Sequence 17, Application US/09132553  
 Patent No. 6183960  
 GENERAL INFORMATION:  
 APPLICANT: Lizardi, Paul M.  
 TITLE OF INVENTION: Rolling Circle Replication Reporter Systems  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS: Patent  
 ADDRESS: 2800 One Atlantic Center  
 STREET: 1201 West Peachtree Street

CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30306-3450  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/132,553  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/563,912  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pabst, Patrea L.  
 REGISTRATION NUMBER: 31,284  
 REFERENCE/DOCKET NUMBER: YU115  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404)873-8794  
 TELEFAX: (404)873-8794  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 29 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-09-132-553-17

Query Match 0.5%; Score 22; DB 4; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 4708 tccatcaaaaaaaaaaaaaa 4729  
 DB 22 tccatcaaaaaaaaaaaaaa 1

RESULT 7  
 US-09-831-570-2/C  
 Sequence 2, Application US/08831570  
 Patent No. 5959175  
 GENERAL INFORMATION:  
 APPLICANT: Thomas, Terry L.  
 APPLICANT: Numborg, Andrew D.  
 TITLE OF INVENTION: A NUCLEAR ARBITRARY 5' REGULATORY REGION  
 TITLE OF INVENTION: FOR THE MODIFICATION OF PLANT SEED LIPID  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Scully, Scott, Murphy & Presser  
 STREET: 400 Garden City Plaza  
 CITY: Garden City  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 11530  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/831,570  
 FILING DATE: 09-Apr-1997  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:

NAME: Digilio, Frank S.  
 REGISTRATION NUMBER: 31,346  
 REFERENCE/DOCKET NUMBER: 10545  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (516)44-4343  
 TELEFAX: (516)44-4343  
 INFORMATION FOR SEQ ID NO: 62:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 843 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 Molecule type: DNA (genomic)  
 US-08-811-570-2

Query Match 0.5%: Score 22; DB 3; Length 843;  
 Best Local Similarity 100.0%: Pred. No. 7.9;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatacaaaaaaaaaaaaaa 4730  
 Db 306 CCAATCAAAAAAAAAAAAAA 285

RESULT 8  
 US-09-248-335-35  
 Sequence 35, Application US/09248335  
 Patent No. 6096504  
 GENERAL INFORMATION:  
 APPLICANT: O'KEEFE, DANIEL  
 TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES  
 FILE REFERENCE: CL-1128-A  
 CURRENT APPLICATION NUMBER: US/09/248,335  
 EARLIER FILING DATE: 1999-02-10  
 EARLIER APPLICATION NUMBER: 08/924,759  
 NUMBER OF SEQ ID NOS: 74  
 SOFTWARE: MICROSOFT Word Version 7.0A  
 SEQ ID NO 35  
 LENGTH: 1098  
 TYPE: DNA  
 ORGANISM: maize  
 US-09-248-335-35

Query Match 0.5%: Score 22; DB 3; Length 1098;  
 Best Local Similarity 100.0%: Pred. No. 7.7;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatacaaaaaaaaaaaaaa 4730  
 Db 1005 ccatcaaaaaaaaaaaaaa 1026

RESULT 9  
 US-08-991-426-1  
 Sequence 1, Application US/08991426  
 Patent No. 6013257  
 GENERAL INFORMATION:  
 APPLICANT: Pan, Yang  
 TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR  
 NUMBER OF SEQ ID NOS: 18  
 CORRESPONDENCE ADDRESS: 18  
 ADDRESSER: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows95  
 SOFTWARE: MICROSOFT Word Version 2.0  
 CURRENT APPLICATION DATA: Windows Version 2.0  
 APPLICATION NUMBER: US/08/991,426  
 FILING DATE: 16-DEC-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/851,160  
 FILING DATE: 05-MAY-1997  
 PRIOR APPLICATION DATA: 08/543,798  
 FILING DATE: 07-MAY-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Melk, John, Ph.D., Anita L.  
 REGISTRATION NUMBER: 35,283  
 REFERENCE/DOCKET NUMBER: 09404/304001  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1784 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 Molecule type: cDNA  
 NAME/KEY: Coding Sequence  
 LOCATION: 40...1224  
 US-08-991-426-1

Query Match 0.5%: Score 22; DB 3; Length 1784;  
 Best Local Similarity 100.0%: Pred. No. 7.4;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatacaaaaaaaaaaaaaa 4730  
 Db 1750 CCAATCAAAAAAAAAAAAAA 1771

RESULT 10  
 US-09-143-470-1  
 Sequence 1, Application US/09143470  
 Patent No. 6043086  
 GENERAL INFORMATION:  
 APPLICANT: Pan, Yang  
 TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR  
 FILE REFERENCE: 09404/049001  
 CURRENT APPLICATION NUMBER: US/09/143,470  
 EARLIER FILING DATE: 1998-08-28  
 NUMBER OF SEQ ID NOS: 20  
 SOFTWARE: PARSED FOR Windows Version 3.0  
 SEQ ID NO 1  
 LENGTH: 1784  
 TYPE: DNA  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: CDS  
 POSITIONS: 1(40)...(1224)  
 US-09-143-470-1

Query Match 0.5%: Score 22; DB 3; Length 1784;  
 Best Local Similarity 100.0%: Pred. No. 7.4;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4709 ccatacaaaaaaaaaaaaaa 4730  
 Db 1750 CCAATCAAAAAAAAAAAAAA 1771

```

RESULT 11
US-08-126-587C-8
Sequence 8, Application US/08126587C
Patent No. 5830562
GENERAL INFORMATION:
APPLICANT: Hayden, Michael
APPLICANT: Goldberg, Paul
APPLICANT: Andrew, Susan
APPLICANT: Rommens, Johanna M.
APPLICANT: Liu, Biacyang
TITLE OF INVENTION: Process for Isolating Genes and the Gene
TITLE OF INVENTION: Polyadenylation of the Gene
TITLE OF INVENTION: Polyadenylation of the Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 1211 E. Morehead Street
CITY: Charlotte
STATE: No. 553438Ch Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/126,587C
FILING DATE: June 21, 1996
CLASSIFICATION: 4,35
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 28,607
REFERENCE/DOCKET NUMBER: 3477-84
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704-377-1561
TELEFAX: 704-377-2034
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4032 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
APPLICANT: Leder, Philip
TITLE OF INVENTION: Detection and Treatment of Breast
NUMBER OF SEQUENCES: 14
ADDRESS: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2223
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.1)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,247B
FILING DATE: August 11, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bieker-Brady, Kristina
REGISTRATION NUMBER: 39,109

```

```

ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,542B
FILING DATE: June 21, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 42840-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TEXT:
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 4032 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-619-542B-28
Query Match 0.43; Score 21; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 4710 calcaaaaaaaaaaaaaaaaaa 4730
Db 211 CATCAAAAAAAAAAAAAAAAAA 231
RESULT 13
US-08-289-247B-1
Sequence 1, Application US/08289247B
Patent No. 5728579
GENERAL INFORMATION:
APPLICANT: Morrison, Br1998 W.
APPLICANT: Leder, Philip
TITLE OF INVENTION: Detection and Treatment of Breast
NUMBER OF SEQUENCES: 14
ADDRESS: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2223
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,247B
FILING DATE: August 11, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bieker-Brady, Kristina
REGISTRATION NUMBER: 39,109

```

REFERENCE/DOCKET NUMBER: 00383/021001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 428-0200  
TELEFAX: (617) 428-7045  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 522  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-289-2478-1

Query Match  
Best Local Similarity 100.0%; Score 21; DB 1; Length 522;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4710 cctcaaaaaaaaaaaaaa 4730  
DB 499 CATCAAAAAAAAAAAAAA 519

RESULT 14  
PCT-US95-09762-1  
SEQUENCE 1, Application PC/TUS9509762  
GENERAL INFORMATION:  
APPLICANT: Morrison, Br199s W.  
ATTORNEY/AGENT INFORMATION:  
TITLE OF INVENTION: Detection and Treatment of Breast  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
ZIP: 02110-3204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
PCT APPLICATION NUMBER: PCT/US95/09762  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/289,247  
FILING DATE: August 11, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00383/021001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-8906  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 522  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US95-09762-1

Query Match  
Best Local Similarity 0.4%; Score 21; DB 5; Length 522;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4710 cttcaaaaaaaaaaaaaa 4730

DB 499 CATCAAAAAAAAAAAAAA 519

RESULT 15  
US-09-004-731-9  
SEQUENCE 9, Application US/09004731  
PATENT NO. 6177258  
GENERAL INFORMATION:  
APPLICANT: Wu Hunter, Shirley  
ATTORNEY/AGENT INFORMATION:  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
PCT APPLICATION NUMBER: US/09/004,731  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,699  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Corneli, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 08/218-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURES:  
NAME/KEY: CDS  
LOCATION: 1..695  
NAME/KEY: Xaa - any amino acid  
LOCATION: 187  
US-09-004-731-9

Query Match  
Best Local Similarity 0.4%; Score 21; DB 4; Length 775;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4710 cttcaaaaaaaaaaaaaa 4730  
DB 754 CATCAAAAAAAAAAAAAA 774

RESULT 16  
US-09-004-731-11/C  
SEQUENCE 11, Application US/09004731  
PATENT NO. 6177258  
GENERAL INFORMATION:  
APPLICANT: Wu Hunter, Shirley  
ATTORNEY/AGENT INFORMATION:  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
PCT APPLICATION NUMBER: US/09/004,731  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,699  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Corneli, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 08/218-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURES:  
NAME/KEY: CDS  
LOCATION: 1..695  
NAME/KEY: Xaa - any amino acid  
LOCATION: 187  
US-09-004-731-9



TITLE OF INVENTION: ELFA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF INVENTIONS: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,731  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
FILING DATE: US/08/749,699  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-004-731-11

Query Match 0.4%: Score 21; DB 4; Length 775;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4710 catcaaaaaaaaaaaaaa 4730  
DB 22 CATCAAAAAAAAAAAAAA 2

RESULT 17  
US-08-749-699-9  
Sequence 9, Application US/08749699  
Patent No. 6210920  
GENERAL INFORMATION:  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Stiegler, Gary  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 103  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,699  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
FILING DATE: US/08-749-699-11/C  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE: CDS  
NAME/KEY: Xaa - any amino acid  
LOCATION: 187  
US-08-749-699-9

FILING DATE: 424  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE: CDS  
NAME/KEY: Xaa - any amino acid  
LOCATION: 187  
US-08-749-699-9

Query Match 0.4%: Score 21; DB 4; Length 775;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4710 catcaaaaaaaaaaaaaa 4730  
DB 754 CATCAAAAAAAAAAAAAA 774

RESULT 18  
US-08-749-699-11/C  
Sequence 11, Application US/08749699  
Patent No. 6210920  
GENERAL INFORMATION:  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Stiegler, Gary  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

```

:
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-749-699-11

Query Match
Best Local Similarity 100.0% Pred. No. 20:
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4710 catcaaaaaaaaaaaaaa 4730
DB 22 CATCAAAAAAAAAAAAAA 2

RESULT 19
US-08-702-344-18
: Sequence 18, Application US/08702344
: Patent No. 5723315
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavallee, Edward
: APPLICANT: Labele, David
: APPLICANT: Mott, Maurice
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: STREET: 87 Cambridge Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: COMPUTER: IBM compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/702,344
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME:
: REGISTRATION NUMBER: A 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1053 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-702-344-18

Query Match
Best Local Similarity 100.0% Pred. No. 19:
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4710 catcaaaaaaaaaaaaaa 4730
DB 1031 CATCAAAAAAAAAAAAAA 1051

RESULT 20
US-08-725-532A-2/G
: Sequence 21, Application US/08725532A

: Patent No. 6020179
: GENERAL INFORMATION:
: APPLICANT: Gail, Surya K.
: TITLE OF INVENTION: NOVEL HUMAN TYROSINE PHOSPHATASES
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: U.S.
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTED Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/725,532A
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0130 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-451-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1275 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: LIBRARY:
: CLONE: Consensus
US-08-725-532A-2

Query Match
Best Local Similarity 100.0% Pred. No. 17:
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4710 catcaaaaaaaaaaaaaa 4730
DB 40 CATCAAAAAAAAAAAAAA 20

RESULT 21
US-08-707-399E-1
: Sequence 1, Application US/08707399E
: Patent No. 6008014
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan and Gimeno, Carlos
: TITLE OF INVENTION: Lipid Metabolic Pathway Compositions
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSER: LAHIVE & COCKFIELD, LLP
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/707,399E
FILING DATE: September 4, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoutas
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: AMT-006
TELEPHONE: (617)227-7400
TELEFAX: (617)227-7400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 114..1130
US-08-707-399E-1

Query Match          0.4%: Score 21; DB 3; Length 1368;
Best Local Similarity 100.0%: Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4710 catcaaaaaaaaaaaaaa 4730
DB 1337 CATCAAAAAAAAAAAAAA 1357

RESULT 22
US-08-052-205-2
Sequence 2, Application US/08052205
Patent No. 5,80759
GENERAL INFORMATION:
APPLICANT: SUGAMURA, KAZUO
APPLICANT: TAKESHITA, TOSHIKAZU
APPLICANT: ASANO, HIROMICHI
APPLICANT: NAKAMURA, MASATKA
APPLICANT: SHIMAMURA, TOSHIRO
APPLICANT: SUZUKI, YASUHIRO
APPLICANT: HAMURO, JUNJI
TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/052,205
FILING DATE: 06-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 104947/1992
FILING DATE: 23-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: OBOLON, NO. 551025man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-615-0X

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1393 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-052-205-2

Query Match          0.4%: Score 21; DB 1; Length 1393;
Best Local Similarity 100.0%: Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4710 catcaaaaaaaaaaaaaa 4730
DB 1371 CATCAAAAAAAAAAAAAA 1391

RESULT 23
US-08-595-974-2
Sequence 2, Application US/08595974
Patent No. 5705608
GENERAL INFORMATION:
APPLICANT: SUGAMURA, KAZUO
APPLICANT: TAKESHITA, TOSHIKAZU
APPLICANT: ASANO, HIROMICHI
APPLICANT: NAKAMURA, MASATKA
APPLICANT: SHIMAMURA, TOSHIRO
APPLICANT: SUZUKI, YASUHIRO
APPLICANT: HAMURO, JUNJI
TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/595,974
FILING DATE: 06-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,205
FILING DATE: 22-APR-1993
APPLICATION NUMBER: JP 104947/1992
FILING DATE: 23-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: OBOLON, NO. 5705608man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-615-0X
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1393 base pairs
TYPE: nucleic acid

```

STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-595-974-2

Query Match 0.4% Score 21: DB 1: Length 1393;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
Y 4710 cctcaaaaaaaaaaaaaa 4730  
Db 1371 cctcaaaaaaaaaaaaaa 1391

## RESULT 24

US-08-052-205-3  
Sequence 3, Application US/08052205  
Patent No. 5510259  
GENERAL INFORMATION:  
APPLICANT: SUGAMURA, KAZUO  
APPLICANT: TAKESHITA, TOSHIKAZU  
APPLICANT: ASAO, HIRONOBU  
APPLICANT: NAKAMURA, MASATAKA  
APPLICANT: SHIMAMURA, TOSHIRO  
APPLICANT: SUZUKI, MANABU  
APPLICANT: HANUKU, JUNJI  
TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/052, 205  
FILING DATE: 1993/04/22  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 104947/1992  
FILING DATE: 23-Apr-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 551025man F.  
REGISTRATION NUMBER: 24,618  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-3240  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1470 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 15..1121  
NAME/KEY: mol\_peptide

LOCATION: 81..1121  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 15..80  
US-08-052-205-3

Query Match 0.4% Score 21: DB 1: Length 1470;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
Y 4710 cctcaaaaaaaaaaaaaa 4730  
Db 1448 cctcaaaaaaaaaaaaaa 1468

## RESULT 25

US-08-595-974-3  
Sequence 3, Application US/08595974  
Patent No. 5705608  
GENERAL INFORMATION:  
APPLICANT: SUGAMURA, KAZUO  
APPLICANT: TAKESHITA, TOSHIKAZU  
APPLICANT: ASAO, HIRONOBU  
APPLICANT: NAKAMURA, MASATAKA  
APPLICANT: SHIMAMURA, TOSHIRO  
APPLICANT: SUZUKI, MANABU  
APPLICANT: HANUKU, JUNJI  
TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/595, 974  
FILING DATE: 06-FEB-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,205  
FILING DATE: 22-Apr-1993  
APPLICATION NUMBER: JP 104947/1992  
FILING DATE: 23-Apr-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 551025man F.  
REGISTRATION NUMBER: 24,618  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1470 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 15..1121  
NAME/KEY: mol\_peptide

NAME/REV: mat.peptide  
LOCATION: 81.1121  
FEATURE: s19.peptide  
NAME/REV: s19.peptide  
LOCATION: 15..80  
US-08-595-974-3

Query Match 0.4% Score 21; DB 1; Length 1470;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 4710 catcaaaaaaaaaaaaaa 4730  
Db 1448 catcaaaaaaaaaaaaaa 1468

RESULT 26  
US-08-872-302-1  
Sequence 1, Application US/08872302  
Patent No. 5846784  
GENERAL INFORMATION:  
APPLICANT: Hitz, William D  
TITLE OF INVENTION: Fatty Acid Modifying Enzymes From  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. duPont de Nemours and Co.  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: USA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT FILING DATE: 1998-08-05  
APPLICATION NUMBER: US/08/872.302  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Majarian, William R  
REGISTRATION NUMBER: P-41,173  
REFERENCE/DOCKET NUMBER: BB-1084  
TELEPHONE: 302-492-4926  
TELEFAX: 302-773-0154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1476 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOL. WEIGHT: 1000000  
FEATURE: type: CDS  
NAME/KEY: CDS  
LOCATION: 134..1279  
US-08-872-302-1

Query Match 0.4% Score 21; DB 2; Length 1476;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4710 catcaaaaaaaaaaaaaa 4730  
Db 1455 catcaaaaaaaaaaaaaa 1475

RESULT 27  
US-09-108-020-11

Sequence 11, Application US/09108020A  
Patent No. 6143561  
GENERAL INFORMATION:  
APPLICANT: Randell, Mark L.  
APPLICANT: Johnston, Mark L.  
APPLICANT: Miernyk, Jan A.  
APPLICANT: Luethy, Michael B.  
APPLICANT: Mooney, Brian P.  
TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND  
TITLE OF INVENTION: ENHANCED CHAIN OXALIC DEHYDROGENASE COMPONENTS TO  
TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANATE BIOSYNTHESIS IN PLANTS  
FILE REFERENCE: UMO 1482  
CURRENT APPLICATION NUMBER: US/09/108.020A  
EARLIER FILING DATE: 1998-06-30  
EARLIER APPLICATION NUMBER: 60/051,291  
EARLIER FILING DATE: 1997-06-30  
EARLIER APPLICATION NUMBER: 60/055,255  
EARLIER FILING DATE: 1998-08-05  
EARLIER APPLICATION NUMBER: 60/076,544  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 11  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-108-020-11

Query Match 0.4% Score 21; DB 3; Length 1587;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 4710 catcaaaaaaaaaaaaaa 4730  
Db 1554 catcaaaaaaaaaaaaaa 1574

RESULT 28  
US-09-360-197-13  
Sequence 13, Application US/09360197  
Patent No. 6287859  
GENERAL INFORMATION:  
APPLICANT: Bassilada, Frederic  
APPLICANT: Bassilada, Frederic  
APPLICANT: Waldmann, Rainer  
APPLICANT: Dwellig, Jan R.  
TITLE OF INVENTION: Human and Rat Families of Neuronal Acid-Sensitive  
FILE REFERENCE: 989,6706P  
CURRENT APPLICATION NUMBER: US/09/360.197  
CURRENT FILING DATE: 1998-08-05  
CURRENT APPLICATION NUMBER: 60/095,408  
PRIOR FILING DATE: 1998-08-05  
PRIOR APPLICATION NUMBER: 60/095,408  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 13  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-360-197-13

Query Match 0.4% Score 21; DB 4; Length 1736;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4710 catcaaaaaaaaaaaaaa 4730  
Db 1711 catcaaaaaaaaaaaaaa 1731

RESULT 29  
 US-07-695-564-4  
 Sequence 4, Application US/07695564  
 Patent No. 5589570  
 GENERAL INFORMATION:  
 APPLICANT: Tamura, Richard N.  
 APPLICANT: Quaranta, Vito  
 TITLE OF INVENTION: INTEGRIN ALPHA SUBUNIT CYTOPLASMIC  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESS:  
 STREET: 11300 Sorrento Valley Road, Suite 200  
 CITY: San Diego  
 STATE: California  
 COUNTRY: United States  
 ZIP: 92121  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 SOFTWARE: PC-DOS/MS-DOS  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/695-564  
 FILING DATE: 19910503  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Filling, Thomas  
 REGISTRATION/DOCKET NUMBER: 34,163  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-546-1555  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5499 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 NAME/KEY: misc.feature  
 LOCATION: 1..5499  
 OTHER INFORMATION: /product= "Human ALPHA 6B"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 1..3260  
 OTHER INFORMATION: /note= "The sequence of SEQ ID NO:4  
 OTHER INFORMATION: is identical to SEQ ID NO:2 between nucleotides 1  
 FEATURE: and 3260."  
 NAME/KEY: misc.feature  
 LOCATION: 3261..5499  
 OTHER INFORMATION: /note= "Nucleotides 3261-5499 of  
 OTHER INFORMATION: SEQ ID NO:4 are identical to nucleotides 3391-5629  
 OTHER INFORMATION: of SEQ ID NO:2. SEQ ID NO:4 has a 130 nucleotide  
 OTHER INFORMATION: deletion in relation to SEQ ID NO:2."  
 NAME/KEY: misc.feature  
 LOCATION: 2921..3325  
 OTHER INFORMATION: /note= "Encompasses the sequence of  
 OTHER INFORMATION: the ALPHA 6B cDNA amplified using primers  
 OTHER INFORMATION: US-07-695-564-4  
 OTHER INFORMATION: 1156/1157."

Query Match 0.4%, Score 21: DB 1: Length 5499;  
 D Local Similarity 100.0%; Fred. No. 17;  
 Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4710 catcaaaaaaaaaaaaaa 4730  
 |||||||  
 Db 5458 CATCAAAAAAAAAAAAAA 5478  
 RESULT 30  
 US-08-241-387-4  
 Sequence 4, Application US/08241387  
 Patent No. 5589570  
 GENERAL INFORMATION:  
 APPLICANT: Tamura, Richard N.  
 APPLICANT: Quaranta, Vito  
 TITLE OF INVENTION: INTEGRIN ALPHA SUBUNIT CYTOPLASMIC  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESS:  
 STREET: The Scripps Research Institute  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: US  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 SOFTWARE: PC-DOS/MS-DOS  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/241,387  
 FILING DATE: 10-MAY-1994  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/695,564  
 FILING DATE: 03-MAY-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Filling, Thomas  
 REGISTRATION/DOCKET NUMBER: 34,163  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-554-2937  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5499 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 NAME/KEY: misc.feature  
 LOCATION: 1..5499  
 OTHER INFORMATION: /product= "Human ALPHA 6B"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 1..3260  
 OTHER INFORMATION: /note= "The sequence of SEQ ID NO:4  
 OTHER INFORMATION: is identical to SEQ ID NO:2 between nucleotides 1  
 FEATURE: and 3260."  
 NAME/KEY: misc.feature  
 LOCATION: 3261..5499  
 OTHER INFORMATION: /note= "Nucleotides 3261-5499 of  
 OTHER INFORMATION: SEQ ID NO:4 are identical to nucleotides 3391-5629  
 OTHER INFORMATION: of SEQ ID NO:2. SEQ ID NO:4 has a 130 nucleotide  
 OTHER INFORMATION: deletion in relation to SEQ ID NO:2."  
 NAME/KEY: misc.feature  
 LOCATION: 2921..3325  
 OTHER INFORMATION: /note= "Encompasses the sequence of  
 OTHER INFORMATION: the ALPHA 6B cDNA amplified using primers  
 OTHER INFORMATION: US-07-695-564-4  
 OTHER INFORMATION: 1156/1157."



```

Query Match      0.48; Score 21; DB 1; Length 5629;
Best Local Similarity 100.00; P-Id. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4710 CATTCAAAAAAAAAAAAAA 4730
      |||
Db 5588 CATTCAAAAAAAAAAAAAA 5608

```

OTHER INFORMATION: Description of artificial sequence: primer p41  
US-09-306-290-37

Db 23 ATCAAAAAAAAAAAAAA 4

OTHER INFORMATION: n - A, T, C or G  
US-08-688-988-19

OY 4711 atcaaaaaaaaaaaaaaa 4730

```

; ATTORNEY/AGENT INFORMATION
; NAME: Connell, Gary J.

```



REGISTRATION NUMBER: 32,020  
 REFERENCE/DOCKET NUMBER: 2618-43  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (303) 863-0223  
 TELEFAX: (303) 863-0223  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 282 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FNAME/REV: CDS  
 LOCATION: 1..103  
 US-08-558-735-10

Query Match 0.4%; Score 20; DB 1; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4711 atcaaaaaaaaaaaaaaa 4730  
 ||||||||||||||||||  
 DB 263 ATCAAAAAAAAAAAAAAA 282

RESULT 36  
 US-08-906-480-10  
 Sequence 25; Application US/08906480  
 Patent No. 6207158  
 GENERAL INFORMATION:  
 APPLICANT: Tripp, Cynthia A.  
 APPLICANT: Brandt, Kevin S.  
 APPLICANT: Wisniewski, Nancy  
 TITLE OF INVENTION: PARASITIC HELMINTH MACROPHAGE MIGRATION  
 TITLE OF INVENTION: INHIBITOR FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sheridan Ross & McIntosh  
 STREET: 1700 Lincoln Street, Suite 3500  
 CITY: Denver  
 STATE: Colorado  
 COUNTRY: U.S.A.  
 ZIP: 80203  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/906,480  
 FILING DATE:  
 CLASSIFICATION:  
 PENDING APPLICATION DATA:  
 APPLICATION NUMBER: 08/558,735  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Connell, Gary J.  
 REGISTRATION NUMBER: 32,020  
 REFERENCE/DOCKET NUMBER: 2618-43  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (303) 863-0223  
 TELEFAX: (303) 863-0223  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 282 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FNAME/REV: CDS  
 LOCATION: 1..103

NAME/REV: CDS  
 LOCATION: 1..103  
 US-08-906-480-10

Query Match 0.4%; Score 20; DB 4; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4711 atcaaaaaaaaaaaaaaa 4730  
 ||||||||||||||||||  
 DB 263 ATCAAAAAAAAAAAAAAA 282

RESULT 37  
 US-08-664-5968-25  
 Sequence 25; Application US/086645968  
 Patent No. 6587012  
 GENERAL INFORMATION:  
 APPLICANT: Jacobs, Kenneth  
 APPLICANT: McCoy, John  
 APPLICANT: Lavallee, Edward  
 APPLICANT: Racie, Lisa  
 APPLICANT: Merberg, David  
 APPLICANT: Treacy, Maurice  
 APPLICANT: Spaulding, Vikki  
 APPLICANT: Bowman, Michael  
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
 TITLE OF INVENTION: ENCODING THEM  
 NUMBER OF SEQUENCES: 37  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 Landings Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/664,5968  
 FILING DATE:  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brown, Scott A.  
 REGISTRATION NUMBER: 32,724  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 452-8124  
 TELEFAX: (617) 452-8124  
 INFORMATION FOR SEQ ID NO: 25:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 371 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-664-5968-25

Query Match 0.4%; Score 20; DB 1; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4711 atcaaaaaaaaaaaaaaa 4730  
 ||||||||||||||||||  
 DB 351 ATCAAAAAAAAAAAAAAA 370

RESULT 38

```

US-08-739-775-3
? Sequence 3, Application US/08739775
? Patent No. 5837490
? GENERAL INFORMATION:
? APPLICANT: Jacobs, Kenneth
? APPLICANT: McCoy, John
? APPLICANT: McCoy, John
? APPLICANT: Reagle, Lisa
? APPLICANT: Merberg, David
? APPLICANT: Treacy, Maurice
? APPLICANT: Spaulding, Vikki
? TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
? NUMBER OF INVENTION: ENCODING THEM
? CORRESPONDENCE ADDRESS:
? ADDRESS: Genetics Institute, Inc.
? STREET: 87 Cambridgepark Drive
? CITY: Cambridge
? STATE: Massachusetts
? COUNTRY: U.S.A.
? ZIP: 02140
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/739,775
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Brown, Scott A.
? REGISTRATION NUMBER: 32,724
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 498-8224
? TELEFAX: (617) 876-5851
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 371 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
US-08-739-775-3

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 371;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4711 atcnaaaaaaaaaaaaaa 4730
DB 351 atcnaaaaaaaaaaaaaa 370

RESULT 39
US-08-558-735-1
? Sequence 1, Application US/08558735
? Patent No. 5681724
? GENERAL INFORMATION:
? APPLICANT: Trlpp, Cynthia A.
? APPLICANT: Brandt, Kevin S.
? APPLICANT: Wisniewski, Nancy
? TITLE OF INVENTION: PARASITIC HELMINTH MACROPHAGE MIGRATION
? NUMBER OF INVENTION: INHIBITOR FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND
? CORRESPONDENCE ADDRESS:
? ADDRESS: Sheridan Ross & McIntosh
? STREET: 1700 Lincoln Street, Suite 3500
? CITY: Denver
? STATE: Colorado
? COUNTRY: U.S.A.

```

```

ZIP: 80203
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/558,735
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Cornell, Gary J.
? REGISTRATION NUMBER: 32,020
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (303) 863-9700
? TELEFAX: (303) 863-0223
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 532 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 8...355
US-08-558-735-1

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 532;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4711 atcnaaaaaaaaaaaaaa 4730
DB 513 atcnaaaaaaaaaaaaaa 532

RESULT 40
US-08-906-480-1
? Sequence 1, Application US/08906480
? Patent No. 6011580
? GENERAL INFORMATION:
? APPLICANT: Trlpp, Cynthia A.
? APPLICANT: Brandt, Kevin S.
? APPLICANT: Wisniewski, Nancy
? TITLE OF INVENTION: PARASITIC HELMINTH MACROPHAGE MIGRATION
? NUMBER OF INVENTION: INHIBITOR FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND
? CORRESPONDENCE ADDRESS:
? ADDRESS: Sheridan Ross & McIntosh
? STREET: 1700 Lincoln Street, Suite 3500
? CITY: Denver
? STATE: Colorado
? COUNTRY: U.S.A.
? ZIP: 80203
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/906,480
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/558,735
? ATTORNEY/AGENT INFORMATION:
? NAME: Cornell, Gary J.
? REGISTRATION NUMBER: 32,020

```

REFERENCE/DOCKET NUMBER: 2618-43  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 base pairs  
TYPE: double  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 8..355  
US-08-906-480-1

Query Match 0.48: Score 20: DB 4: Length 532:  
Best Local Similarity 100.0%: Pred. No. 50:  
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 4711 atcaaaaaaaaaaaaaa 4730  
Db 513 ATCAAAAAAAAAAAAAA 532

RESULT 41  
US-08-539-304A-1  
Sequence 1: Application US/08539304A  
Patent No. 5792933  
GENERAL INFORMATION:  
APPLICANT: MA, DIN-POW  
TITLE OF INVENTION: FIBER-SPECIFIC PROTEIN EXPRESSION IN THE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
STREET: 1755 JEFFERSON DAVIS HWY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Patented Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/539,304A  
FILING DATE: 04-OCT-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: NORMAN, ORLON F  
REGISTRATION NUMBER: 24,618  
PRIORITY/COINVENTOR NUMBER: 21,513-037-27  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 84..443  
US-08-539-304A-1

Query Match 0.48: Score 20: DB 1: Length 615:  
Best Local Similarity 100.0%: Pred. No. 49:

Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 4711 atcaaaaaaaaaaaaaa 4730  
Db 595 ATCAAAAAAAAAAAAAA 614

RESULT 42  
US-08-451-405A-2  
Sequence 2: Application US/08451405A  
Patent No. 5736358  
GENERAL INFORMATION:  
APPLICANT: FASEL, NICOLAS JOSEPH  
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE  
TITLE OF INVENTION: DICTYOSTELIUM EXPRESSION VECTOR AND  
METHOD FOR EXPRESSING A DESIRED PROTEIN  
CORRESPONDENCE ADDRESS:  
STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE  
CITY: PITTSBURGH  
STATE: PENNSYLVANIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 15219-1818  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 FLOPPY DISK  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: WORDPERFECT 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,405A  
FILING DATE: 26-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,273  
PRIORITY NUMBER: 07/965,199  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: UNKNOWN  
US-08-451-405A-2

Query Match 0.48: Score 20: DB 1: Length 731:  
Best Local Similarity 100.0%: Pred. No. 49:  
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 4711 atcaaaaaaaaaaaaaa 4730  
Db 295 ATCAAAAAAAAAAAAAA 314

RESULT 43  
US-07-912-900-23  
Sequence 23: Application US/07912900  
Patent No. 5349125  
GENERAL INFORMATION:  
APPLICANT: Holton, Timothy A.  
APPLICANT: Cornish, Edwin C.  
APPLICANT: Kovacic, Filipa  
APPLICANT: Tanaka, Toshihazu  
APPLICANT: Kishimoto, Akihiro  
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: U.S.A.

```

? ZIP: 11530
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? SOFTWARE: PC-DOS/MS-DOS
? CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
? APPLICATION NUMBER: US/07/912,900
? FILING DATE: 19920713
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: DIGILLO, Frank S.
? REGISTRATION NUMBER: 31,346
? REFERENCE/DOCKET NUMBER: 86332
? TELEPHONE: (516) 742-4343
? TELEFAX: (516) 742-4366
? TELETYPE: 230 901 SANS UR
? INFORMATION FOR SEQ ID NO: 23:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 733 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-07-912-900-23

```

```

Query Match 0.4%; Score 20; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

? Oy 4711 atcagagagagagagagagagag 4730
? Db 714 ATCAAAAAAAAAAAAAAAAAAAAA 733

```

```

RESULT 44
US-08-285-309-23
? Sequence 27 Application US/08285109
? Patent No. 5568832
? GENERAL INFORMATION:
? APPLICANT: Holton, Timothy A.
? APPLICANT: Cornish, Edwin C.
? APPLICANT: Kovacic, Filippa
? APPLICANT: Tanaka, Yoshikazu
? APPLICANT: Lester, Diane R.
? TITLE OF INVENTION: GENETIC SEQUENCES ENCODING A 3.5'-
? NUMBER OF SEQUENCES: 23
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Scully, Scott, Murphy & Plessner
? STREET: 400 Garden City Plaza
? CITY: Garden City
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 11530
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? SOFTWARE: PC-DOS/MS-DOS
? CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
? APPLICATION NUMBER: US/08/285,309
? FILING DATE: 03-AUG-1994
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: DIGILLO, Frank S.
? REGISTRATION NUMBER: 31,346
? REFERENCE/DOCKET NUMBER: 86332
? TELEPHONE: (516) 742-4343
? TELEFAX: (516) 742-4366

```

```

? TELEX: 230 901 SANS UR
? INFORMATION FOR SEQ ID NO: 23:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 733 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-285-309-23

```

```

Query Match 0.4%; Score 20; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

? Oy 4711 atcagagagagagagagagagag 4730
? Db 714 ATCAAAAAAAAAAAAAAAAAAAAA 733

```

```

RESULT 45
US-08-313-075A-27
? Sequence 27 Application US/08313075A
? Patent No. 5568832
? GENERAL INFORMATION:
? APPLICANT: Holton, Timothy A.
? APPLICANT: Cornish, Edwin C.
? APPLICANT: Tanaka, Yoshikazu
? TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
? NUMBER OF SEQUENCES: 58
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Scully, Scott, Murphy & Plessner
? STREET: 400 Garden City Plaza
? CITY: Garden City
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 11530
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? SOFTWARE: PC-DOS/MS-DOS
? CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
? APPLICATION NUMBER: US/08/313,075A
? FILING DATE: 30-NOV-1994
? CLASSIFICATION: 800
? PRIOR APPLICATION DATA:
? PRIOR APPLICATION NUMBER: AU PL 1538/92
? FILING DATE: 27-MAR-1992
? PRIOR APPLICATION DATA:
? PRIOR APPLICATION NUMBER: AU PL 6698/93
? FILING DATE: 07-JAN-1993
? APPLICATION NUMBER: AU PCT/AU93/00127
? FILING DATE: 25-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: DIGILLO, Frank S.
? REGISTRATION NUMBER: 31,346
? REFERENCE/DOCKET NUMBER: 86332
? TELEPHONE: (516) 742-4343
? TELEFAX: (516) 742-4366
? TELETYPE: 230 901 SANS UR
? INFORMATION FOR SEQ ID NO: 27:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 733 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA
? NAME/KEY: CDS

```

• Thu Apr 11 07:44:00 2002

LOCATION: 1..402  
US-08-313-075A-27

```

Query Match      0.4%: Score 20; DB 1; Length 733;
Best Local Similarity 100.0%: Pred No. 49;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4711 atcaaaaaaaaaaaaaaaaaa 4730
    |||
Db 714 ATCAAAAAAAAAAAAAAAAAA 733

```

Search completed: April 10, 2002, 18:10:10  
Job time: 7444 sec

**THIS PAGE BLANK (USPTO)**









|        |                                   |        |                              |
|--------|-----------------------------------|--------|------------------------------|
| 9837   | 10705: contig of 869 bp in length | 42068  | 42933: gap of unknown length |
| 10706  | gap of unknown length             | 42934  | contig of 866 bp in length   |
| 11579: | contig of 874 bp in length        | 43801  | contig of 868 bp in length   |
| 11580  | gap of unknown length             | 44675: | contig of 874 bp in length   |
| 12459: | contig of 860 bp in length        | 45338: | contig of 863 bp in length   |
| 12460  | contig of 863 bp in length        | 45339  | gap of unknown length        |
| 13321: | contig of 865 bp in length        | 46405: | contig of 867 bp in length   |
| 13322  | gap of unknown length             | 47269: | contig of 861 bp in length   |
| 14190: | contig of 869 bp in length        | 47270  | gap of unknown length        |
| 14191  | gap of unknown length             | 48139: | contig of 870 bp in length   |
| 15058: | contig of 868 bp in length        | 48999: | contig of 860 bp in length   |
| 15059  | gap of unknown length             | 49000  | gap of unknown length        |
| 15923: | contig of 865 bp in length        | 49867: | contig of 867 bp in length   |
| 15924  | gap of unknown length             | 50729: | gap of unknown length        |
| 16802: | contig of 872 bp in length        | 51588: | contig of 859 bp in length   |
| 16803  | gap of unknown length             | 52460: | contig of 872 bp in length   |
| 17680: | contig of 878 bp in length        | 53323: | contig of 863 bp in length   |
| 17681  | gap of unknown length             | 54189: | contig of 866 bp in length   |
| 18548: | contig of 868 bp in length        | 55053: | contig of 864 bp in length   |
| 18549  | gap of unknown length             | 55910: | contig of 857 bp in length   |
| 19413: | contig of 865 bp in length        | 55911  | gap of unknown length        |
| 20289: | gap of unknown length             | 56776: | contig of 866 bp in length   |
| 20290  | contig of 867 bp in length        | 57669: | contig of 893 bp in length   |
| 21156: | contig of 867 bp in length        | 58529: | contig of 860 bp in length   |
| 21157  | gap of unknown length             | 59404: | contig of 875 bp in length   |
| 22031: | contig of 875 bp in length        | 60263: | contig of 859 bp in length   |
| 22032  | gap of unknown length             | 60264  | gap of unknown length        |
| 22916: | contig of 885 bp in length        | 61140: | contig of 877 bp in length   |
| 22917  | gap of unknown length             | 61141  | gap of unknown length        |
| 23786: | contig of 870 bp in length        | 61997: | gap of unknown length        |
| 23787  | gap of unknown length             | 62885: | contig of 888 bp in length   |
| 24655: | contig of 869 bp in length        | 63756: | contig of 871 bp in length   |
| 24656  | gap of unknown length             | 64624: | contig of 868 bp in length   |
| 25519: | contig of 864 bp in length        | 64625  | gap of unknown length        |
| 25520  | gap of unknown length             | 65487: | gap of unknown length        |
| 26383: | contig of 864 bp in length        | 65488  | contig of 859 bp in length   |
| 27253: | contig of 870 bp in length        | 67207: | contig of 861 bp in length   |
| 27254  | gap of unknown length             |        |                              |
| 28119: | contig of 864 bp in length        |        |                              |
| 28120  | gap of unknown length             |        |                              |
| 28987: | contig of 868 bp in length        |        |                              |
| 28988  | gap of unknown length             |        |                              |
| 29861: | contig of 874 bp in length        |        |                              |
| 29862  | gap of unknown length             |        |                              |
| 30737: | contig of 876 bp in length        |        |                              |
| 30738  | gap of unknown length             |        |                              |
| 31601: | contig of 864 bp in length        |        |                              |
| 31602  | gap of unknown length             |        |                              |
| 32481: | contig of 880 bp in length        |        |                              |
| 32482  | gap of unknown length             |        |                              |
| 33347: | contig of 866 bp in length        |        |                              |
| 33348  | gap of unknown length             |        |                              |
| 34218: | contig of 871 bp in length        |        |                              |
| 34219  | gap of unknown length             |        |                              |
| 35081: | contig of 873 bp in length        |        |                              |
| 35082  | gap of unknown length             |        |                              |
| 35948: | contig of 867 bp in length        |        |                              |
| 35949  | gap of unknown length             |        |                              |
| 36814: | contig of 866 bp in length        |        |                              |
| 36815  | gap of unknown length             |        |                              |
| 37687: | contig of 873 bp in length        |        |                              |
| 37688  | gap of unknown length             |        |                              |
| 38583: | contig of 896 bp in length        |        |                              |
| 38584  | gap of unknown length             |        |                              |
| 39452: | contig of 869 bp in length        |        |                              |
| 39453  | gap of unknown length             |        |                              |
| 40334: | contig of 882 bp in length        |        |                              |
| 40335  | gap of unknown length             |        |                              |
| 41201: | contig of 867 bp in length        |        |                              |
| 42067: | gap of unknown length             |        |                              |
| 42067: | contig of 866 bp in length        |        |                              |

## alignment\_scores:

Quality: 75.00 Length: 75  
 Ratio: 1.000  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-750-590-2 x AC022466/rev

Align seg 1/1 to reverse of: AC022466 from: 1 to: 83775

440 SeqAlasGlaaPaRcGtLsLcGcInaGtLcLcAaHtLsYgV 456  
 |||||||  
 21494 TCAGCAAAACAGACGACGTCAAGCTCAAAATATACGTGCGACCAAGG 21445



seq\_name: gb\_htg:AC010076

```
seq_documentation_block:
  locs: AC010076 155760 bp
  cna: 1
```

SEQUENCE, 19 ordered pieces

**KEYWORDS** HTG; HTGS\_PHASE2; HTGS\_DRAFT

1010

## REFERENCE AUTHORS

JOURNAL  
REFERENCE  
AUTHORS

**JOURNAL  
COMMENT**

b.hcg:AC010076

Location block: 55760 bp      HPG      O1-SEP-2000  
Homo sapiens      Human genome clone KRL1-64K10, WORKING DRAFT  
AC010076      AC010076.5 GI:95977994  
SEQUENCE: 19 ordered pieces.  
HTG: HTGS\_PHA82; HTGS\_DRAFT.

Homo sapiens  
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 153760)  
Roven L., Madan A., Qin S., Baradarni L., Birditt B., Bloom S.,  
Dicks R., Fleetwood P., Harrison G., Kaur A., Madan A.,  
Nelson J., Nierlich D. and Wood L.  
Sequencing of human chromosome 15 D1S114-D1S115 region  
Unpublished

2 (bases 1 to 153760)  
Roven L., Madan A., Qin S., Abbadai N., Baradarni L., Birditt B.,  
Bloom S., Dicks R., Fleetwood P., Harrison G., James R., Laaky S.,  
Lorenson T., Madan A., Nelson J., Nierlich D., Shaffer T. and Wood L.

Submitted (11-SEP-1999) Multibase Sequencing Center, University  
of Washington, PO Box 357370, Seattle, WA 98195 USA  
On Sep 1, 2000 this sequence version replaced sf:812412.

----- Genome Center -----

Center: Multimegabase Sequencing Center  
Center code: UNMSC  
Web site: [http://chroma.mbt.washington.edu/msg\\_www](http://chroma.mbt.washington.edu/msg_www)  
Contact: [leerowens@systemsbiology.org](mailto:leerowens@systemsbiology.org)  
Summary Statistics

Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 90% of reads  
Chemistry: Dye-primer Big Dye; 10% of reads  
Assembly program: Phrap; version 0.990359  
Insert size: 150000; agarose-gel

**Sequence Quality Assessment:** This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. Quality scores for manually edited bases have been reduced to quality zero. Quality scores in 10,000 bp windows are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the genome browser file format but are available as part of this entry's ASN.1 file.

NOTE: This is a working draft sequence. It currently consists of 19 contigs. Gaps between the contigs is believed to be as large as 800 bp. The order of the pieces of the gaps between them are based on estimates that have been provided by the author.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

| 1     | 39368: contig of 39368 bp in length |
|-------|-------------------------------------|
| 33669 | 39468: gap of unknown length        |
| 33670 | 41356: gap of unknown length        |
| 43857 | 44356: contig of 4448 bp in length  |
| 44057 | 52135: gap of unknown length        |
| 52336 | 54693: gap of unknown length        |
| 52336 | 54693: contig of 2457 bp in length  |
| 54693 | 58291: gap of unknown length        |
| 54793 | 58291: contig of 3499 bp in length  |
| 58392 | 61782: gap of unknown length        |
| 58392 | 61782: contig of 3391 bp in length  |
| 61783 | 61882: gap of unknown length        |
| 61883 | 75899: contig of 14017 bp in length |

| FEATURES            | source                               |
|---------------------|--------------------------------------|
| 75960               | 75959: gap of unknown length         |
| 75961               | 75960: contig of 1954 bp in length   |
| 75962               | 75961: gap of unknown length         |
| 76034               | 89130: gap of unknown length         |
| 89130               | 89289: gap of unknown length         |
| 89290               | 95543: contig of 6254 bp in length   |
| 95543               | 95634: gap of unknown length         |
| 95634               | 95702: contig of 2259 bp in length   |
| 95644               | 95702: gap of unknown length         |
| 95645               | 101823: gap of unknown length        |
| 96003               | 101925: gap of unknown length        |
| 101826              | 101925: gap of unknown length        |
| 101925              | 107779: contig of 5854 bp in length  |
| 107780              | 107879: gap of unknown length        |
| 107880              | 133178: contig of 15299 bp in length |
| 13279               | 13278: gap of unknown length         |
| 13279               | 13278: gap of unknown length         |
| 132811              | 132921: gap of unknown length        |
| 132921              | 133029: contig of 3819 bp in length  |
| 132921              | 133129: gap of unknown length        |
| 133130              | 133965: contig of 6636 bp in length  |
| 133965              | 13966: gap of unknown length         |
| 13966               | 151596: contig of 1631 bp in length  |
| 151596              | 151596: gap of unknown length        |
| 151597              | 151596: contig of 15154 bp in length |
| 151597              | 151596: contig of 15154 bp in length |
| Location/Qualifiers |                                      |
| 1..155760           |                                      |

|            | clone_11b | RPCI | human | BAC | library | 11     |
|------------|-----------|------|-------|-----|---------|--------|
| BASE COUNT | 45760     | a    | 33275 | c   | 32273   | g      |
| ORIGIN     |           |      |       |     | 42185   | t      |
|            |           |      |       |     |         | 2267   |
|            |           |      |       |     |         | others |

```

alignment_scores:
  Quality: 75.00
  Ratio: 1.000
  Gaps: 0
Percent Similarity: 100.000
Percent Identity: 100.000

```

05\*03-750-590-2 X AC010076/rev ..

TCAGCAATCAGACCGACTGAAGCTCCAAATGAAGCTGGCACACAAGT 134845

134844 GGCAGATGCAAGCTTTAGCATTAGAATGTGAAGGGTCAACGAGGATT 134795

134794 CAGATGACAGATTAAGCATTAGAGATGCATTAAAGATGTGCAGAG 134745

134744 AGGATGTATGAGTCAGAAAGGTAAGTTAAACAATGCAGACCACTTTCT 134695

134694 TGCCCTTAAGACACTTAACAAGT 134670

|                          |                                         |             |
|--------------------------|-----------------------------------------|-------------|
| seq_documentation_block: |                                         |             |
| LOCUS                    | CEFC3VS                                 | 4643 bp     |
| DEFINITION               | Canis familiaris mRNA for C3VS protein. | MAN         |
| ACCESSION                | X99145                                  |             |
| VERSION                  | X99145.2                                | GI:10944717 |
| KEYWORDS                 | C3VS protein.                           |             |
| C3VS protein.            |                                         |             |
|                          |                                         | 20-OCT-2000 |



```

151 spval1apgl1yagthrpoleuval1leu1al1eth1g1lmet 164
|||||
145 ATGTAGACGGCGGACACCACTTGTCTGCTACCTACGATG 185

seq_name: gb_pat:AX071358
seq_documentation_block: 401 bp DNA PAT 25-JAN-2001
LOCUS AX071358
DEFINITION Sequence 1830 from Patent W00102568.
ACCESSION AX071358
VERSION AX071358.1 GI:12581709
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
MIMAT: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 401)
Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J.,
Kasam, A., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D.,
Lambert, I., Leshkovitz, D., Kita, D., Garcia, V. and Strache, C., B.
Human genes and gene expression products.
CHIRON CORP. (US) : HSE0, INC. (US)
FEATURES
location/Qualifiers
1..401
/organism="Homo sapiens"
/db_xref="taxon:9606"
mlec_feature
1..401
/Note="A,T,C or G"
BASE COUNT 114 a 85 c 107 g 93 t 2 others
ORIGIN

alignment_scores:
Quality: 47.00 Length: 47
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x AX071358

Align seg 1/1 to: AX071358 from: 1 to: 401

118 Aspleu1ng1yagthral1eul1asap1al1al1eul1aspcysr 134
|||||
45 GACCTGACGGAGAACACCTGACCTGACGACGACGACGATGTC 94
134 oSer1ieg1leu1eul1eul1asap1al1al1eul1asap1al1eul1a 151
|||||
95 TTCTAGCATACACGCTCTTGTGACCATGGCGGCTCTGTGATCCAAAG 144

151 spval1apgl1yagthrpoleuval1leu1al1eth1g1lmet 164
|||||
145 ATGTAGACGGCGGACACCACTTGTCTGCTACCTACGATG 185

seq_name: gb_pat:AX193084
seq_documentation_block: 441 bp DNA PAT 15-AUG-2001
LOCUS AX193084
DEFINITION Sequence 651 from Patent W00149716.
ACCESSION AX193084
VERSION AX193084.1 GI:15211039
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
MIMAT: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 441)
Xu, J., Lodes, M.J., Secrist, H., Benson, D.R., Meagher, M.J.,
Stolk, J.A., King, G.E., Wang, T. and Jiang, Y.
Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use

```

```

JOURNAL Patent: WO 0149716-A 651 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
location/Qualifiers
1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 127 a 97 c 114 g 102 t 1 others
ORIGIN

alignment_scores:
Quality: 47.00 Length: 47
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x AX193084

Align seg 1/1 to: AX193084 from: 1 to: 441

118 Aspleu1ng1yagthral1eul1asap1al1al1eul1aspcysr 134
|||||
137 GACCTGACGGAGAACACCTGACCTGACGACGACGACGATGTC 186
134 oSer1ieg1leu1eul1eul1asap1al1al1eul1asap1al1eul1a 151
|||||
187 TTCTAGCATACACGCTCTTGTGACCATGGCGGCTCTGTGATCCAAAG 236
151 spval1apgl1yagthrpoleuval1leu1al1eth1g1lmet 164
|||||
237 ATGTAGACGGCGGACACCACTTGTCTGCTACCTACGATG 277

seq_name: gb_pat:AX192628
seq_documentation_block: 421 bp DNA PAT 15-AUG-2001
LOCUS AX192628
DEFINITION Sequence 195 from Patent W00149716.
ACCESSION AX192628
VERSION AX192628.1 GI:15210591
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
MIMAT: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 421)
Secrist, H., Benson, D.R., Meagher, M.J.,
Stolk, J.A., King, G.E., Wang, T. and Jiang, Y.
Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
Patent: WO 0149716-A 195 12-JUL-2001.
CORIXA CORPORATION (US)
FEATURES
location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 109 a 101 c 125 g 86 t
ORIGIN

alignment_scores:
Quality: 37.00 Length: 37
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x AX192628

Align seg 1/1 to: AX192628 from: 1 to: 421

64 Ser1yag1yagthrpoleuval1leu1al1eth1g1lmet 80
|||||
310 TCAGACGGAGAACACCTGACCTGACGACGACGACGATGTC 359

```



```

repeat_region      /rpl_family="12", complement(18147..18235)
repeat_region      /rpl_family="12MAA"
repeat_region      18351..18509
repeat_region      /rpl_family="MERSA"
repeat_region      18510..18625
repeat_region      /rpl_family="MR3"
repeat_region      complement(18811..19056)
repeat_region      /rpl_family="MR"
repeat_region      complement(19213..19160)
repeat_region      19794..19947
repeat_region      /rpl_family="MR"
repeat_region      complement(20054..20111)
repeat_region      /rpl_family="MERSB"
repeat_region      20112..20420
repeat_region      /rpl_family="Alusg"
repeat_region      20431..20714
repeat_region      /rpl_family="AluGo"
repeat_region      complement(20715..20798)
repeat_region      /rpl_family="MERSB"
repeat_region      20799..20968
repeat_region      /rpl_family="Alusg/X"
repeat_region      20972..21283
repeat_region      /rpl_family="Alusg"
repeat_region      complement(21287..21327)
repeat_region      21357..21485
repeat_region      /rpl_family="MERSB"
repeat_region      21822..21943
repeat_region      /rpl_family="MR"
repeat_region      21944..22247
repeat_region      /rpl_family="Alusx"
repeat_region      22248..22311
repeat_region      /rpl_family="MR"
repeat_region      22356..22376
repeat_region      /rpl_family="AT-rich"
repeat_region      23070..23383
repeat_region      /rpl_family="Alusg"
repeat_region      23685..23976
repeat_region      /rpl_family="AluY"

alignment_scores:
    Quality: 36.00
    Ratio: 1.000
    Percent Similarity: 100.000
    Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x AC009369 ..

Align seg 1/1 to: AC009369 from: 1 to: 172850

285 ASPLANDLILLEGUASNGUASPLRYSGLUAGUEAIGYSLIGI 301
171344 GATTTCGAGATTGAAAGCAAGATTTTAAACAGAGTTGAGAAATATCA 171393
301 nctngtctgactatgctatdeuauasplrysvaiaendcylleuglndleuglnl 318
171394 GCAGAACACAGAAACATCTTGGATCAAAAGTACATGCTTACAGTACAGC 171443
318 euasngcu 320
|||||||
171444 TGAATGAG 171451

seq_name: qb_sts:G35947
seq_documentation_block:
LOCUS      418 bp      mRNA
DEFINITION  SMC-5623 human homo sapiens STS cDNA, sequence tagged
ACCESSION  G35947.1 GI:2734614
KEYWORDS   STS.
SOURCE     human.
```



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Myers R.M.  
 TITLE Human STS (1997)  
 JOURNAL Unpublished (1997)  
 COMMENT

Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Department of Genetics, N-344, Stanford, CA 94305, USA  
 Phone: 415/2596688  
 Fax: 415/2596689  
 Email: myers@shgc.stanford.edu  
 Primer A: ATGCTCCGACGCGCTACG  
 Primer B: GGACAGGTTGACGCTCTG  
 STS size: 167  
 PCR profile:  
 Initial incubation: 95 degrees C for 10 minutes  
 Denaturation: 94 degrees C for 30 seconds  
 Annealing: 60 degrees C for 30 seconds  
 Polymerization: 72 degrees C for 23 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9600

Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 Amplify: each 200 uM  
 Amplifying Gold Polymerase: 0.07 units/uL  
 Total Vol: 5 uL

Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

Prepared with primer pairs derived from M69830 -- Unigene.  
 Location/Qualifiers

FEATURES  
 source 1..418  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="151b-Human"  
 /clone="151b-Human"  
 SFS  
 primer\_bind complement(298..317)  
 BASE COUNT 83 a 92 c 65 g 178 t  
 ORIGIN

alignment\_scores: 25.00 Length: 25  
 Overall: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x G36947/rev ..

Align seq 1/1 to reverse of: G36947 from: 1 to: 418

1156 lvscluglulphcglulysglulwaglytletlcyalaasetleuagcl 1172

140 AAGAGACATTGGAGAAAGAGCTTGATCATTAAGACCGACTGAGAGA 91

1172 lylscluglulserglnasnllys 1180

90 AAGAGAGAGAAAGCAAAACAA 66

seq\_name: gb\_hlg:AC022466

seq documentation block: 83775 bp DNA HTG 04-FEB-2000  
 LOCUS AC022466 Homo sapiens chromosome 15 clone CTD-2027K34 map 15q23, LOW-PASS  
 DEFINITION  
 SEQUENCE SAMPLING.

AC022466  
 AC022466.1 GI:689623  
 HTG: HTGS\_PHASED.  
 SOURCE  
 human.

ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Myers R.M.  
 TITLE Human STS (1997)  
 JOURNAL Unpublished (1997)  
 AUTHORS Myers R.M.

Sequencing of human chromosome 15 D15S146-D15S17 region  
 Unpublished (1997)  
 Submitted (04-FEB-2000) Multimegabase Sequencing Center, University  
 of Washington, PO BOX 357720, Seattle, WA 98195, USA  
 James R., Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T.  
 and Hood L.  
 Direct Submission  
 Submitted (04-FEB-2000) Multimegabase Sequencing Center, University  
 of Washington, PO BOX 357720, Seattle, WA 98195, USA  
 James R., Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T.  
 and Hood L.  
 Center: Multimegabase Sequencing Center  
 Center code: UMNSC  
 Web site: http://chroma.mbl.washington.edu/msg.www  
 Contact: leecrowen@u.washington.edu

Summary Statistics  
 Sequencing vector: pUC18: 108752  
 Chemistry: Dye-terminator Big Dye, 90% of reads  
 Chemistry: Dye-terminator Big Dye, 10% of reads  
 Assembly program: Phrap: Version 0.990399

NOTE: This record contains 96 individual  
 sequencing reads that have not been assembled into  
 contigs. Runs of N are used to separate the reads  
 and the order in which they appear is completely  
 arbitrary. Low-pass sequence sampling is useful for  
 identifying clones that may be deleted or  
 rearranged. It should not be assumed that this clone  
 will be sequenced to completion. In the event that  
 the record is updated, the accession number will  
 be preserved.

1 887: contig of 887 bp in length  
 888 1780: contig of 893 bp in length  
 1781 2645: contig of 865 bp in length  
 2646 3525: contig of 880 bp in length  
 3526 4401: contig of 876 bp in length  
 4402 5290: contig of 889 bp in length  
 5291 6155: contig of 853 bp in length  
 6156 7152: contig of 996 bp in length  
 7153 8036: contig of 885 bp in length  
 8037 8948: contig of 912 bp in length  
 8949 9836: gap of unknown length  
 9837 10705: contig of 869 bp in length  
 10706 11579: contig of 874 bp in length

|       |                                   |       |                                   |
|-------|-----------------------------------|-------|-----------------------------------|
| 11580 | 12459: contig of 880 bp in length | 42934 | 43801: contig of 868 bp in length |
| 12460 | 13321: contig of 862 bp in length | 43802 | 44675: contig of 874 bp in length |
| 13322 | 14130: contig of 863 bp in length | 44676 | 45538: contig of 863 bp in length |
| 14191 | 15058: contig of 868 bp in length | 45539 | 46405: contig of 867 bp in length |
| 15059 | 15923: contig of 865 bp in length | 46406 | 47269: contig of 864 bp in length |
| 15924 | 16802: contig of 879 bp in length | 47270 | 48139: contig of 863 bp in length |
| 16803 | 17680: contig of 872 bp in length | 48140 | 48999: contig of 860 bp in length |
| 17681 | 18548: contig of 868 bp in length | 49000 | 49866: contig of 867 bp in length |
| 18549 | 19413: contig of 865 bp in length | 49867 | 50729: contig of 863 bp in length |
| 19414 | 20289: contig of 876 bp in length | 50730 | 51588: contig of 859 bp in length |
| 20290 | 21156: contig of 867 bp in length | 51589 | 52460: contig of 872 bp in length |
| 21157 | 22031: contig of 875 bp in length | 52461 | 53323: contig of 863 bp in length |
| 22032 | 22916: contig of 885 bp in length | 53324 | 54189: contig of 866 bp in length |
| 22917 | 23786: contig of 870 bp in length | 54190 | 55053: contig of 864 bp in length |
| 23787 | 24655: contig of 869 bp in length | 55054 | 55910: contig of 857 bp in length |
| 24656 | 25519: contig of 865 bp in length | 55911 | 56776: contig of 866 bp in length |
| 25520 | 26383: contig of 864 bp in length | 56777 | 57669: contig of 893 bp in length |
| 26384 | 27253: contig of 870 bp in length | 57670 | 58529: contig of 860 bp in length |
| 27254 | 28119: contig of 866 bp in length | 58530 | 59404: contig of 870 bp in length |
| 28120 | 28987: contig of 868 bp in length | 59405 | 60263: contig of 859 bp in length |
| 28988 | 29861: contig of 874 bp in length | 60264 | 61141: contig of 877 bp in length |
| 29862 | 30737: contig of 876 bp in length | 61141 | 61997: contig of 857 bp in length |
| 30738 | 31601: contig of 864 bp in length | 61998 | 62885: contig of 868 bp in length |
| 31602 | 32481: contig of 880 bp in length | 62886 | 63756: contig of 871 bp in length |
| 32482 | 33347: contig of 866 bp in length | 63757 | 64624: contig of 868 bp in length |
| 33348 | 34218: contig of 871 bp in length | 64625 | 65487: contig of 863 bp in length |
| 34219 | 35081: contig of 863 bp in length | 65488 | 66346: contig of 859 bp in length |
| 35082 | 35948: contig of 867 bp in length | 66347 | 67207: contig of 861 bp in length |
| 35949 | 36814: contig of 866 bp in length |       |                                   |
| 36815 | 37687: contig of 873 bp in length |       |                                   |
| 37688 | 38583: contig of 896 bp in length |       |                                   |
| 38584 | 39452: contig of 869 bp in length |       |                                   |
| 39453 | 40334: contig of 863 bp in length |       |                                   |
| 40335 | 41201: contig of 867 bp in length |       |                                   |
| 41202 | 42067: contig of 866 bp in length |       |                                   |
| 42068 | 42933: contig of 866 bp in length |       |                                   |

## alignment scores:

Quality: 20.00 Length: 20  
 Percent Similarity: 100.000 Gaps: 0  
 Percent Identity: 100.000

## alignment block:

US-09-750-590-2 x AC022466

Align seg 1/1 to: AC022466 from: 1 to: 83775

1284 Thrlucluglunrlygargllleclnglunseralalygslolleglunlaly 1300

6 ACAGAGTTTCACAGAGACATACAGAGATCTGCTTAACCAATATAGACAAA 55

1300 AAGAAATAG 55

56 ACATATATAG 55

```

seq_name: gb_pat:AR003316
LOCUS       AR003316             60 bp            DNA                    PAT
DEFINITION   Sequence 38 from patent US 5744300.
ACCESSION    AR003316
VERSION      AR003316.1 GI:3964575
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 60)
AUTHORS      Linkens,M.H.K., Hirsch,K.S., Villepontoux,B., Feng,J., Funk,M. and
TITLE        Methods and reagents for the identification and regulation of
JOURNAL      senseless-related genes
PATENT       Patent: US 5744300-A, 38-28-Apr-1998;
              Location/Qualifiers
                source          1..60
                               /organism="Unknown"
BASE COUNT   27 a         4 c         17 g         12 t
ORIGIN
alignment_scores:
Quality:      19.00           Length:      19
Ratio:        1.000          Gaps:       0
Percent Similarity: 100.000   Percent Identity: 100.000
Alignment Block:
US-03-750-590-2 x AR003316 ..
Align seg 1/1 to: AR003316 from: 1 to: 60
477 Tlctygtgntcgcctatgaagtcacggcgccgcaaggcatgctatgca 493
|||||
3 ATAAAGCAATATTGAAGAATCCATTAAAGATGTCGACAGACGATGTATCA 52
493 usertgu 495
|||||
53 CTCMAA 59
seq_name: gb_pr:AK000990
seq_documentation_block:
LOCUS       AK000990             1628 bp            mRNA                    PRI
DEFINITION   Homo sapiens CDNA FJ10128 f1s, clone HBMB1002997, weakly similar
ACCESSION    AK000990
VERSION      AK000990.1 GI:7021992
KEYWORDS     oligo capping; f1s (full insert sequence);
              homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
              mRNA, clone_11b:HBMB1 clone:HBMB1002997.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
(1 bases) Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosohiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakami,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Salto,K., Yamamoto,J., Wakamatsu,A.,
Nekamura,Y., Nagahira,T., Masuko,T., Niimiya,K. and Iwayanagi,T.
NEO human cDNA sequencing project
Reference seq. ID: FJ10128
REFERENCE    1 (bases 1 to 1628)
AUTHORS      Isogai,T. and Otsuki,T.
TITLE        Direct Submission
JOURNAL
Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Heili research institute, Genomics laboratory; 1532-3 Yana
Kitsurazu, Chiba 297-0812, Japan (e-mail:genom@centri.co.jp).

```

```

COMMENT
    P01-81-438-52-3951, Fax:81-438-52-3952)
    NEDO human cDNA sequencing project supported by Ministry of
    International Trade and Industry of Japan; cDNA full insert
    sequencing; Research Association for Biotechnology; DNA library
    construction; 3' end analysis; PCR amplification; DNA selection;
    etc.) The institution supported by Japan Key Technology Center
    and Department of Virology, Institute of Medical Science,
    University of Tokyo.

FEATURES
    Source
        1..1628
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="HEMMA100297"
        /clone_1lb="HEMMA1"
        /tissue="embryo; 10 weeks"
        /isline_type="whole embryo, mainly head"
        /note="cloning vector: pME18SPJ3"
        22..>1628
        /note="unnamed protein product"
        /codon_start=1
        /protein_id="BAA91457.1"
        /db_xref="GI:7021993"
        /translation="MKRHKSLSEIROLKREINRYKAKLQHYVPEPYSOYSELKD
        KNSLALDGLNLLKLLKLKKLLEKLLKLLNNITLKNNHYPLAYSDSKK
        LKEDLDRLDLDTOKYTEKILDEMLLLENDLSLDSQISLTLPVPPEPKESIK
        IIAKSNVYEALKDELSEKKKGDEOKIHATLSSENTLKKMSNOIVPYTEHVKKK
        LINDTLKRNELLEDYKKKEFDINOEPYAIDKNLETLRNLQNMIOVAETYSIAEHL
        EKMSISLOSRRKQVDSONAEIIANRCGCEIYLTHAIATAKKEILOICIKVYKH
        PVSEKSSFKFATEKRLADLSEDTQGVSVSEEVKKKEDNKILKEIFLRKDI
        RORTYLECSHMBRALSRKTDLINKOLDISQTEKVWVKEXLEENAGVSTSL
        VOLLQGHVPLEDRALKSLNGTENGLEISRKSNOCYERDQVTVLQLLENOR
        NSSVALERDLQTKAREGEVIKKKS"

BASE COUNT      707 a   236 g   346 g   337 t
ORIGIN

Alignment_scores:
    Quality: 17.00          Length: 17
    Ratio: 1.000           Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

Alignment_block:
    US-09-750-550-2 x AK000930 ..

Align seg 1/1 to: AK000930 from: 1 to: 1628

1069 TyscLgslsLeawAlGicUluUsmlAtyGcIntHrSeGcUlUleuaU 1085
|||||
1321 AAACAGACTGTAGTGAACAAATGCCAAAAGACTTCCTCAGATTCC 1370

1085 a 1085 .
1371 A 1371

seq_name: gb.sts:c14728

seq_documentation_block:
LOCUS      G14728                250 bp            STS             04-JAN-1996
DEFINITION human STS SHGC-134459.
ACCESSION  G14728
VERSION    G14728.1 GI:1130467
KEYWORDS
SOURCE     human
ORGANISM  Homo sapiens
REFERENCE  1 [bases 1 to 250]
AUTHORS   Myers,R.M.
JOURNAL   Unpublished (1995)
COMMENT
Contact: Richard M. Myers

```

Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Fax: 4157259689  
 Email: myers@shgc.stanford.edu  
 Primer A: AAGCCCGAAGAACCCG  
 Primer B: TGCATCAGATTCACACG  
 STS size: 148  
 PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds

Annealing: 94 degrees C for 23 seconds

Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Tag Polymerase: 0.05 units/uL

Total Vol: 10 uL

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 20 mM

pH: 8.3

Prepared with primer pairs derived from T47625--Merck/UNLEST.

Location/Qualifiers

1..250

/organism="Homo sapiens"

STS

primer\_bind

22..169

complement(149..169)

BASE COUNT 70 a 63 c 56 g 58 t 3 others

ORIGIN

alignment\_scores:

Quality: 13.00 Length: 13

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x G14728/rev ..

Align seq 1/1 to reverse of: G14728 from: 1 to: 250

Contact: Michael Olivier, David R. Cox

Stanford Human Genome Center

Stanford University School of Medicine

4005 Miranda Ave 2nd Fl., Palo Alto, CA 94025, USA

Tel: (650) 320-5801

Fax: (650) 320-5801

Email: olivier@shgc.stanford.edu

Primer A: TGCCTCTACATCAACATCCTT

Primer B: AATGCTTGCACAGTTCATGCC

STS size: 282

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds

Annealing: 60 degrees C for 30 seconds

Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Amplifed Gold Polymerase: 0.07 units/uL

Total Vol: 5 uL

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 8.3

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed

and developed at the Stanford Human Genome Center.

Location/Qualifiers

1..541

/organism="Homo sapiens"

STS

primer\_bind

225..506

complement(484..506)

BASE COUNT 202 a 97 c 54 g 188 t

ORIGIN

alignment\_scores:

Quality: 11.00 Length: 11

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x G58616/rev ..

Align seq 1/1 to reverse of: G58616 from: 1 to: 541

TITLE Crocker M.C., Hinkle G., Holder M.E. and Sogin M.L.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 688)  
 AUTHORS Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Crocker M.C., Hinkle G., Holder M.E. and Sogin M.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUN-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA  
 COMMENT \* NOTE: This record contains 1 individual  
 \* sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.  
 \* 1  
 \* 688: contig of 688 bp in length.  
 FEATURES  
 source  
 1. 688  
 /organism="Giardia intestinalis"  
 /strain="WB-C6"  
 /db\_xref="taxon:5741"  
 /clone="K1116"  
 /cdate="2000-07-10"  
 /csize="162 g 190 g 143 t 13 others  
 BASE COUNT 180 a 162 c 190 g 143 t 13 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-750-590-2 x AC071902 ..  
 Align seg 1/1 to: AC071902 from: 1 to: 688  
 25 LeuMEcATGAlaAlaAGUAAGCtYAspValGtA 35  
 |||||||  
 339 CTCATCGCGCTCCAGAAAGAGTGACGTACAG 371  
 seq\_name: gb.htg:AC075673  
 seq\_documentation\_block: 932 bp DNA HTG 31-JUL-2000  
 LOCUS AC075673  
 DEFINITION Giardia intestinalis clone EJ7590 strain WB-C6, LOW-PASS SEQUENCE  
 SAMPLING.  
 AC075673.1 GI:9589972  
 VERSION AC075673.1  
 KEYWORDS HTG; HTGS; PHASE0.  
 SOURCE Giardia intestinalis.  
 ORGANISM Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.  
 REFERENCE 1 (bases 1 to 932)  
 Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U., Crocker M.C., Hinkle G., Holder M.E. and Sogin M.L.  
 TITLE Giardia: a model for ancient eukaryotic genome analysis  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 932)  
 AUTHORS Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Crocker M.C., Hinkle G., Holder M.E. and Sogin M.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-JUL-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA  
 COMMENT \* NOTE: This record contains 1 individual  
 \* sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.  
 \* 1  
 \* 932: contig of 932 bp in length.  
 FEATURES  
 source  
 1. 932  
 /organism="Giardia intestinalis"  
 /strain="WB-C6"  
 /db\_xref="taxon:5741"  
 /clone="EJ7590"  
 /cdate="2000-07-10"  
 /csize="255 a 233 c 259 g 185 t  
 BASE COUNT 255 a 233 c 259 g 185 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-750-590-2 x AC075673 ..  
 Align seg 1/1 to: AC075673 from: 1 to: 932  
 25 LeuMEcATGAlaAlaAGUAAGCtYAspValGtA 35  
 |||||||  
 334 CTCATCGCGCTCCAGAAAGAGTGACGTACAG 366  
 seq\_name: gb.htg:AC076402  
 seq\_documentation\_block: 957 bp DNA HTG 01-AUG-2000  
 LOCUS AC076402  
 DEFINITION Giardia intestinalis clone K1116 strain WB-C6, LOW-PASS SEQUENCE  
 SAMPLING.  
 AC076402.1 GI:9624573  
 VERSION AC076402.1  
 KEYWORDS HTG; HTGS; PHASE0.  
 SOURCE Giardia intestinalis.  
 ORGANISM Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.  
 REFERENCE 1 (bases 1 to 957)  
 Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U., Crocker M.C., Hinkle G., Holder M.E. and Sogin M.L.  
 TITLE Giardia: a model for ancient eukaryotic genome analysis  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 957)  
 AUTHORS Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Crocker M.C., Hinkle G., Holder M.E. and Sogin M.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-AUG-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA  
 COMMENT \* NOTE: This record contains 1 individual  
 \* sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.  
 \* 1  
 \* 957: contig of 957 bp in length.  
 FEATURES  
 source  
 1. 957  
 /organism="Giardia intestinalis"  
 /strain="WB-C6"

```

BASE COUNT      198 A    264 C    226 G    264 T    5 others
ORIGIN          /db_xref=taxon:5741*
               /clone="KJ4116"

alignment_scores:
  Quality:      11.00      Length:      11
  Ratio:        1.000      Caps:         0
  Percent Similarity: 100.000  Percent Identity: 100.000

Alignment_block:
US-09-750-590-2 x AC076402/rev ..
Align seg 1/1 to reverse of: AC076402 from: 1 to: 957

25 LeuMetAgaLaiaGuaTgClyAspAlaGlu 35
|||||
545 CTCATCCGGCTCGACGAAGAAGTCGTACGTAGG 513
seq_name: gb_hlg.AC072272

seq_documentation_block:
LOCUS       AC072272             1036 bp            DNA           07-JUN-2000
DEFINITION  Giardia intestinalis clone M4258 strain WB-C6, LOW-PASS SEQUENCE
ACCESSION   AC072272.1 GI:8324620
VERSION     AC072272.1 GI:8324620
KEYWORDS    HG; HTGS_PHASED;
SOURCE      Giardia intestinalis.
ORGANISM    Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
REFERENCE   1 (bases 1 to 1036)
AUTHORS     Lamerdin,J.E., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
            Cockcroft,S.B., Holzer,M.E. and Scott,K.L.:
TITLE       Giardia: a model for ancient eukaryotic genome analysis
JOURNAL     Unpublished
PUBLISHED   2 (bases 1 to 1036)
AUTHORS     Morrison,H.C., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.:
DIRECT SUBMISSION JUN-2000 Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Woods Hole, MA 02543-1015, USA
COMMENT     * NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying regions that may be gene-rich and allows
            * overlap relationships between reads to be detected.
            * However, it should not be assumed that all genes
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source           1..1036; contig of 1036 bp in length.
                     1..1036
                     /organism="Giardia intestinalis"
                     /strain="WB-C6"
                     /db_xref="taxon:5741"
                     /clone="M4258"
BASE COUNT      283 A    244 C    284 G    225 T

alignment_scores:
  Quality:      11.00      Length:      11
  Ratio:        1.000      Caps:         0
  Percent Similarity: 100.000  Percent Identity: 100.000

Alignment_block:
US-09-750-590-2 x AC072272 ..
Align seg 1/1 to: AC072272 from: 1 to: 1036

25 LeuMetAgaLaiaGuaTgClyAspAlaGlu 35
|||||
662 CTCATCCGGCTCGACGAAGAAGTCGTACGTAGG 694
seq_name: gb_pf.AC006930

seq_documentation_block:
LOCUS       AC006930             41055 bp            DNA           PRI            01-MAR-1999
DEFINITION  Homo sapiens chromosome 19, cosmid R3243, complete sequence.
ACCESSION   AC006930
VERSION     AC006930.1 GI:4309680
KEYWORDS    MID
SOURCE      Human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 41055)
AUTHORS     Burkhardt-Schultz,K., Gordon,D., Dias,J., Ramirez,M., Stillwagon,S.,
            Burkhart-Schultz,K., Gordon,D., Dias,J., Ramirez,M., Stillwagon,S.,
            Thompson,L., Beyer,M., Bodmer,W., Regalla,M., Gerry,A., Barnes,J.,
            Luvig,J., Attix,C., Andrejaie,T., Frankham,R., Attick,K., Kellum,J.,
            Coffield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,Q.,
            Kornblatter,B., Ariellano,A., Sanders,C., Ov,d., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carraro,A.V.:
TITLE       Sequence analysis of a 6 kb region in 19p13.3 between CDC34 and
            D19S21 mapped
JOURNAL     Unpublished
PUBLISHED   2 (bases 1 to 41055)
AUTHORS     Lamerdin,J.E.
TITLE       Direct Submission
JOURNAL     National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Cosmid
            map and sequence cosmid R3350 (AC005620) to the left from bases 1
            to 11,988 and cosmid R3350 (AC005620) to the right from bases 33,115
            to 41,055. The cosmid R3243 (AC005718) and cosmid R3243 (AC005718)
            map and sequence information may be obtained at:
            http://www.bio.lnl.gov/bhsr/genome/genome.html.
FEATURES             Location/Qualifiers
     source           1..41055
                     /organism="Homo sapiens"
                     /chromosome="19p13.3"
                     /clone="R3243"
                     /map_19p13.3 between CDC34 and D19S21*
                     /cell_line="SHL2-B"
                     /clone_id="LLN9NC03 R chromosome 19-specific cosmid library"
                     /note="Cosmid library constructed at LML from flow-sorted
                     chromosomes from human-hamster hybrid SHL2-B, which carries
                     chromosome 19 as its only human chromosome."
misc_feature          1..87
                     /note="Blastx similarity to overlapping ESTs:
                     AA890113 at5302.41 Scores: NF-L-T-GBC.SI Homo sapiens cDNA
                     clone IMAGE:1461003.3, similar to contigs PT85.t2 MER25
                     repetitive element ; mRNA sequence [homo sapiens];
                     (102..188); Score: 6.8e-66; 100% identity..AA906192
                     O12805.81 Scores: NF-L-T-GBC.SI Homo sapiens cDNA clone
                     AB9801.1 at5302.41 to R1028599.028597
                     TRANSCRIPTIONAL REGULATORY PROTEIN P12 ANIMALITY.; mRNA
                     sequence [homo sapiens]; Length = 376; (102..188); Score:
                     2.e-107; 100% identity..Length = 376; (102..188); Score:
                     complement(3)in(19)..-243,438..-740,822..-945,1474..-1602,
                     1636..-1731,2873..-2932,4898..-5080,6572..-6699,8264..-8329)"
                     /note="Hypothetical human SIR2-like protein; Replicates
                     protein encoded by R3350.2 (9113548790). Hypothetical
                     protein displays similarity (24%) to the yeast
                     Sir2 protein, a well known silencer SIR2 (encoded by Y06700), and
                     murine Sir2

```

```

/codon_start=1
/evidence=not_experimental
/product="R33423_1"
/protein_id="A015478.1"
/db_xref="GI:4309681"
/translation="MSVAVNAGSLSPVADKQKCGPELIDPEPEELSERVWELARLYWOS
SSVYVHVGASVTSQSDIDFEGHCHVWYHKEGGLAKEDITFESAPLOTMAHVALDLE
NSGRTLVAKKGLKACGCGCPSPDSNOLPHCGEGLDTPLIMDSPLSPRIALADLE
ASRNADLSTTGTSIDTQIPSSGNLPLATKRCGRGLYVYNLQPTSRHADRIAHQYDE
VWRMLRHGLGLEITADVDYVLRALPLPRPPPLPLERKSPESPRNCSIPAPQGE
PCAGHNSPASPAPRRPSPAPAHPEKRPGLVFEENEPTPOF"
/complement(864..945)
/notes="BLASTN similarity to U46339 HSU46339 Human
pancreatic cancer cell line Patu 8988T Homo sapiens cDNA
clone xs346, mRNA sequence [Homo sapiens]; Length = 295;
[214..295]; Score: 1.0e-104; 84% identity."
misc_feature
/complement(1474..1686)
/notes="BLASTN similarity to U46339 HSU46339 Human
pancreatic cancer cell line Patu 8988T Homo sapiens cDNA
clone xs346, mRNA sequence [Homo sapiens]; Length = 295;
[1..213]; Score: 1.0e-104; 99% identity."
misc_feature
repeat_region
2103..2395
/rfp_family="Alusg"
2416..2630
/rfp_family="MTRIC"
3084..3148
/rfp_family="MIR"
/complement(3352..3683)
/rfp_family="Alusg"
/complement(3684..3816)
/rfp_family="Alusg"
/complement(3821..4118)
/rfp_family="Alusg"
4182..4453
/rfp_family="Alusg"
4907..4996
/rfp_family="Alusg"
/notes="BLASTN similarity to overlapping ESTs:
(4907..4996) AA890113.153802.81 Soares.NFL.T.GBC.S1 Homo
sapiens cDNA clone IMAGE:1461003.3 similar to contigs
PTRS.t2 MCR25 repetitive element; mRNA sequence [Homo
sapiens]; (187..276); Score: 6.8e-66; 100%
identity; (4907..4996) AA906192.0386005.81
IMAGE:1506345.3 similar to TR-028597.028597 TRANSCRIPTIONAL REGULATORY
PROTEIN, S1R2 FAMILY; mRNA sequence [Homo sapiens]; Length
= 376; (187..376); Score: 2.8e-107; 100% identity."
misc_feature
/complement(5035..5076)
/notes="BLASTX identity to AA839912 v66a07.r1 Strategene
mouse heart (4937316) Mus musculus cDNA clone
IMAGE:1259892.5', mRNA sequence [Mus musculus];
(138..239); Score: 1.8e-39; 100% identity."
repeat_region
5097..5221
/rfp_family="(GA)n"
5260..5367
/rfp_family="(CACA)n"
/complement(5893..6060)
/rfp_family="Alusg/x"
/complement(6175..6414)
/rfp_family="Alusg"
/notes="BLASTX identity to AA839912 v66a07.r1 Strategene
mouse heart (4937316) Mus musculus cDNA clone
IMAGE:1259892.5', mRNA sequence [Mus musculus]; (69..187);
Score: 1.8e-39; 98% identity."
repeat_region
6831..7145
/rfp_family="LINE2"
7187..7284
/rfp_family="LINE2"

```

```

repeat_region
7295..7598
/rfp_family="Alusx"
/misc_feature
/complement(8264..8329)
/notes="BLASTX identity to AA839912 v66a07.r1 Strategene
mouse heart (4937316) Mus musculus cDNA clone
IMAGE:1259892.5', mRNA sequence [Mus musculus]; (3..65);
Score: 1.8e-39; 100% identity."
repeat_region
/complement(8756..8785)
/rfp_family="(CA)n"
repeat_region
/complement(8786..9069)
/rfp_family="Alusx"
CDS
join(9114..9235,25665..25730,33291..33397,33571..33758,
34354..34591,36035..36162,38360..38488)
/notes="Hypothetical human protein. It is record extend the
hypothetical protein encoding 10887 amino acids. It is
Hypothesized that this protein is involved in the
However, the amino terminus of this protein has not been
identified definitively within this genomic accession, and
will likely require the isolation of cDNA. Displays weak
similarity to ankyrin motif-containing proteins"
/codon_start=3
/evidence=not_experimental
/product="R33423_2, partial protein"
/protein_id="A096482"
/translation="GYLVKVGAGLSDSPKKNHYLSTYVVLGTLSEKGGITVLSODM
GKSDRLIADVENVNDAPRSKATPLLIAMOKCHTDLCELLDQGAANDQDGLGFALM
LAGEGASPEVVEVLQCGKQGITDGLGDAHAGLAKDLILILQEPAAQRPSP
ALVEDSGEASSQSDNDVVEEIVRIJROGRILQKINLEQHEREROERQSLHCS
QVQELQDLIVEROESESIGVESLQSLRSLILE"
/complement(9156..9201)
/rfp_family="LINE2"
9423..9718
/rfp_family="Alusg"
/complement(9769..9873)
/rfp_family="LINE2"
10517..10665
/rfp_family="(TGCAn)"
10679..10862
/rfp_family="(TGCAn)"
10875..10993
/rfp_family="(TGCAn)"
repeat_region
1129..1129
/rfp_family="(TGCAn)"
/complement(11667..12046)
/rfp_family="MIR"
12180..12251
/notes="Predicted exon, program: gtrial2exons_human_1.3,
frame: 2, quality: good, score: 54.000"
repeat_region
/complement(12623..12918)
/rfp_family="LINE2"
12921..13222
/rfp_family="Alusx"
13249..13305
/complement(13249..13305)
/rfp_family="LINE2"
13349..13466
/complement(13365..14166)
/rfp_family="Alusg"
repeat_region
/complement(14172..14469)
/rfp_family="Alusg"
/complement(14597..14899)
/rfp_family="Alusg"
/complement(14902..15029)
/rfp_family="(TAAA)n"

```

```

alignment_scores:
Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
05-03-750-590.2 x AC006930

```

Align seg 1/1 to: AC006930 from: 1 to: 41055

7821 AATGCTCCACCTGCGCCGCAATATACGCGC 27853

seq\_name: gb\_htg:AC027272

seq\_documentation\_block:

LOCUS AC027272 179607 bp DNA HTG 01-SEP-2000

DEFINITION Homo sapiens chromosome 19 clone RP11-486P13, WORKING DRAFT

SEQUENCE 18 unordered pieces.

AC027272.1 GI:9838224

VERSION 1

KEYWORDS HTG; HTGS; HBASE1; HTGS-DRAFT.

SOURCE human

ORGANISM Homo sapiens

Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 179607)

AUTHORS Waterston R.H.

TITLE The sequence of Homo sapiens clone

UNPUBLISHED to 179607)

2 (bases 1 to 179607)

AUTHORS Waterston R.H.

TITLE Direct Submission

UNPUBLISHED

Submitted (28-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

On Aug 17, 2000 this sequence version replaced gi:8570236.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: H.NH0486P13

----- Summary Statistics -----

Sequencing vector: p13: 95%

Chemistry: De-palmer ET: 98% of reads

Assembly program: Phrap: version: 990319

Consensus quality: 170867 bases at least Q40

Consensus quality: 174124 bases at least Q30

Insert size: 188000: agarose-fp

Insert size: 179350: sum-of-contigs

Quality coverage: 4.55 in Q20 bases; sum-of-contigs

Quality coverage: 4.86 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It only consists of 18 contigs. The true order of the contigs is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1338: contig of 1338 bp in length

1339: gap of unknown length

1439: contig of 1892 bp in length

3331: gap of unknown length

3431: contig of 1999 bp in length

5430: gap of unknown length

5529: gap of unknown length

8107: contig of 2578 bp in length

8530: gap of unknown length

8508: gap of unknown length

11380: contig of 2973 bp in length

11181: gap of unknown length

14039: contig of 9755 bp in length

14040: gap of unknown length

14140: contig of 3091 bp in length

17231: gap of unknown length

FEATURES

source

1..179607

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="19"

/clone="RP11-486P13"

/note="assembly\_name:Contig98"

1439..3330

/note="assembly\_name:Contig98"

3431..5429

/note="assembly\_name:Contig10"

5530..8107

/note="assembly\_name:Contig11"

8208..11180

/note="assembly\_name:Contig12"

11281..14039

/note="assembly\_name:Contig13"

14140..17230

/note="assembly\_name:Contig14"

17331..21334

/note="assembly\_name:Contig15

clone\_end:596

vector:1567:left"

25435..32607:right"

26978..31453

/note="assembly\_name:Contig16"

31554..36515

/note="assembly\_name:Contig17"

36516..46678

/note="assembly\_name:Contig18"

46779..57959

/note="assembly\_name:Contig19"

58100..74742

/note="assembly\_name:Contig20"

74843..92504

/note="assembly\_name:Contig21"

92605..119545

/note="assembly\_name:Contig22"

119646..147254

/note="assembly\_name:Contig24"

147255..179607

/note="assembly\_name:Contig25"

BASE COUNT 41107 a 48065 c 48884 g 39812 t 1719 others

ORIGIN

alignment\_scores: quality: 11.00 length: 11

Percent Similarity: 100.000 Gaps: 0

Percent Identity: 100.000











alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-750-590-2 x CELK09F6/rev ..

Align seg 1/1 to reverse of: CELK09F6 from: 1 to: 38737

738 LYSSEHSHASPVALLIEVALASPAAPLKA 747

14379 AATCTCATACGTCACCTGATGATCTGA 14350

seg\_name: gb\_htg:AC078966

seg\_documentation\_block:

AC078966: 42594 bp DNA

DEFINITION Homo sapiens chromosome 8 clone RP11-166H15 map 8, WORKING DRAFT

SEQUENCE 60 unordered pieces.

AC078966.2 GI:12229305

HTG: HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;

Mammalia; Eularchia; Primates; Catarrhini; Hominoidea; Homo.

Blattner, B., Linton, L., Nusbaum, C., and Lander, E.

Unpublished

2 (bases 1 to 142694)

Blattner, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Beda, P., Boguslavsky, L.,

Bouknight, B., Brown, A., Butte, G., Campolongo, A., Castle, A.,

Chepel, T., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

Flitzhugh, M., Gage, D., Gage, J., Gage, S., Gage, S., Gage, S.,

Graham, L., Grand-Pierre, N., Hago, B., Hefford, A., Horton, L.,

Ilav, T., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, M.,

Lamarez, R., Landers, T., Lehoucq, J., Levine, R., Lien, C., Liu, G.,

McDonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K.,

McPherson, R., Melid, J., Menius, L., Miyata, T., Mlewa, Y.,

Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T.,

O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,

Pierre, R., Pisan, C., Pollard, Y., Raymond, C., Ribick, M., Riley, R.,

Soung, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talamas, J., Teebay, S., Theodore, J.,

Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,

Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J.,

Zimmer, A., and Zody, M.

Direct Submission

Submitted (13-ANG-2000) Whitehead Institute/MIT Center for Genome

Research, 330 Charles Street, Cambridge, MA 02141, USA

Research, 330 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A. F. A., & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submission@genome.wi.mit.edu

Center project name: L10521

Center clone name: 166.H.15

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 118069 bases at least Q40

Consensus quality: 129460 bases at least Q50

Consensus quality: 133963 bases at least Q20  
 Insert size: 145000; average-fp  
 Insert size: 136794; sum-of-contigs  
 Quality coverage: 2.9 in Q20 bases; average-fp  
 Quality coverage: 3.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 60 contigs ordered by sequence record id. The gaps between the contigs are represented as arbitrary runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1276: contig of 1276 bp in length  
 1277 1376: gap of 100 bp  
 1278 1377: gap of 100 bp  
 1279 213:2083: contig of 100 bp in length  
 1280 2184:2846: contig of 663 bp in length  
 1281 2847:2946: gap of 100 bp in length  
 1282 2947:3493: contig of 547 bp in length  
 1283 3494:3593: gap of 100 bp  
 1284 3594:3956: contig of 363 bp in length  
 1285 3957:4056: gap of 100 bp  
 1286 4057:4680: contig of 624 bp in length  
 1287 4681:4780: gap of 100 bp in length  
 1288 4781:5728: contig of 100 bp in length  
 1289 5730:5829: contig of 625 bp in length  
 1290 5830:6454: contig of 625 bp in length  
 1291 6455:6554: gap of 100 bp in length  
 1292 6555:7160: contig of 606 bp in length  
 1293 7161:7260: gap of 100 bp  
 1294 7261:7911: contig of 651 bp in length  
 1295 7912:8011: gap of 100 bp  
 1296 8012:8484:8384: contig of 373 bp in length  
 1297 8385:8484:8384: contig of 100 bp in length  
 1298 8485:9700: gap of 100 bp in length  
 1299 9701:10985: contig of 1285 bp in length  
 1300 10986:11085: gap of 100 bp in length  
 1301 11086:12215: contig of 1130 bp in length  
 1302 12216:12315: gap of 100 bp  
 1303 12316:13572: contig of 1257 bp in length  
 1304 13573:13672: gap of 100 bp  
 1305 13673:14638: contig of 966 bp in length  
 1306 14639:14654: gap of 100 bp in length  
 1307 14655:1547: contig of 808 bp in length  
 1308 15478:15647: gap of 100 bp in length  
 1309 15648:17488: contig of 1841 bp in length  
 1310 17489:17588: gap of 100 bp in length  
 1311 17589:18605: contig of 1017 bp in length  
 1312 18606:18705: gap of 100 bp  
 1313 18706:19997: contig of 1292 bp in length  
 1314 19998:20097: gap of 100 bp  
 1315 20098:21511: contig of 100 bp in length  
 1316 21512:23233: contig of 1719 bp in length  
 1317 23234:23333: gap of 100 bp in length  
 1318 23334:24759: contig of 1426 bp in length  
 1319 24760:24859: gap of 100 bp  
 1320 24860:25970: contig of 1111 bp in length  
 1321 25971:26070: gap of 100 bp  
 1322 26071:27476: contig of 1397 bp in length  
 1323 27477:27589: gap of 100 bp in length  
 1324 27590:28968: contig of 1431 bp in length  
 1325 28969:29098: gap of 100 bp in length  
 1326 29099:30232: contig of 1134 bp in length  
 1327 30233:30332: gap of 100 bp in length  
 1328 30333:31691: contig of 1359 bp in length  
 1329 31692:31791: gap of 100 bp in length  
 1330 31792:32790: contig of 999 bp in length  
 1331 32791:32850: gap of 100 bp in length  
 1332 32851:3424: contig of 1334 bp in length



```

Consensus quality: 163088 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 164522; sum-of-continigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-continigs
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 12 continigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the continigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1933 2032: contig of 1932 bp in length
1933 2032: gap of 100 bp
2033 5510: contig of 3478 bp in length
5511 5610: gap of 100 bp
5611 10360: contig of 4750 bp in length
10361 10460: gap of 100 bp
10461 17677: contig of 7217 bp in length
17678 17777: gap of 100 bp
17778 29069: contig of 11292 bp in length
29070 29169: gap of 100 bp
29170 31365: contig of 15205 bp in length
31366 31474: gap of 100 bp
31475 44474: contig of 15072 bp in length
44475 59546: gap of 15072 bp in length
59547 59646: gap of 100 bp in length
59647 76690: contig of 17044 bp in length
76691 76790: gap of 100 bp
76791 94256: contig of 17466 bp in length
94257 94356: gap of 100 bp
94357 115371: contig of 21015 bp in length
115372 115472: gap of 100 bp
115473 136667: contig of 21099 bp in length
136668 165622: gap of 100 bp
165623 28955: contig of 28955 bp in length.
Location/Qualifiers
1. 165622
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="p11-55SEP"
/clone.lib="RC1-11 Human Male BAC"
1. 1932
/note="assembly-fragment"
2033. 5510
/note="assembly-fragment"
5611. 10360
/note="assembly-fragment"
10361. 17677
/note="assembly-fragment"
10361-17677: assembly-fragment
vector_side:right"
17778. 29069
/note="assembly-fragment"
29170. 44374
/note="assembly-fragment"
44475. 59546
/note="assembly-fragment"
59647 76690: assembly-fragment"
76791. 94256
/note="assembly-fragment"
94357. 115371
/note="assembly-fragment"
clone_end:SP6
115472-136667:
115472-136667:
misc_feature
136668. 165622
/note="assembly-fragment"

```

---

```

BASE COUNT      47453 a  40298 c  37118 g  39643 t   1110 others
ORIGIN
alignment_scores:
Quality:      10.00      Length:      10
Ratio:        1.0000     Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment block:
US-09-750-590-2 x AC026950/rev ...
Align seg 1/1 to reverse of: AC026950 from: 1 to: 165622
882 Lysileayaspqluasngtllatlayys 891
|||||
115052 AAAAAAGATGAAAGAAATCTGTAANA 115023
seq_name: gb-pr:CN5072FP
seq_documentation_block:
LOCUS      CN5072FP 192972 bp DNA PRI 13-JUL-2001
DEFINITION Human chromosome 14 DNA sequence BAC R-737F10 of library RC1-11
            from chromosome 14 of Homo sapiens (human), complete sequence.
ACCESSION  AL512360.2 GI:14787751
VERSION    AL512360.2
XREFS      HOMO: HOMO
SOURCE     HTO: HTCS,ACTIVERIN.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 192972)
            Heilig,R., Pett,J.L., Vico,V., Dasilva,C., Robert,C., Winkler,P.,
            Böttler,P., Catolico,L., Barde,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Böckenberg,R., Bruns,T., deBardinas,V., Crouaud,C.,
            Lyspey,G., Saurin,R. and Weissensbach,J.
            Sequencing of the human chromosome 14
            Unpublished
            2 (bases 1 to 192972)
            Genoscope.
            Direct Submission
            Submitted (13-JUL-2001) Genoscope - Centre National de Sequencage ;
            BP 191 91005 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            Web : www.genoscope.cns.fr)
            On Jul 17, 2001 this sequence version replaced g1:11990875.
            Comment: Genoscope / Centre National de Sequencage
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: seqref@genoscope.cns.fr
            -----
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-514D2
            Downstream BAC (overlapping the SP6 end) : C-2082C19 (AC-ALJ58338)
            Assembly program: Phrap, version 2.0
            Quality coverage: 6.58x in Q20 bases; sum-of-continigs
            -----
            Overall quality chart :
            Range      :      bases
            0 - 9      :      5
            10 - 19     :      57
            20 - 29     :      242
            30 - 39     :      1840
            40 - 49     :      4608
            50 - 59     :      6958
            60 - 69     :      20863
            70 - 79     :      44517
            80 - 89     :      103861
            90 - 99     :
            -----

```

FEATURES  
Source  
Location/Qualifiers  
1. 192972  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="R-727P10"  
/clone.lib="Rpci-11"

BASE COUNT 65761 a 34870 c 33542 g 58799 t  
ORIGIN

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

Alignment Block:  
US-09-750-590-2 x CNS07EP9 ..

Align seg 1/1 to: CNS07EP9 from: 1 to: 192972

1018 LysValCysLeuThrLeuGlnGlyAspLeu 1027  
118907 AAGGTCCTTCGTCGACCAACACCTCCTG 118936

seq\_documentation\_block:  
LOCUS AC060775 202993 bp DNA HTG 04-JUL-2001  
DEFINITION Homo sapiens chromosome 8 clone RP11-678D18 map 8, WORKING DRAFT  
SEQUENCE 3 unordered pieces.  
AC060775 2 GI:14595927  
VERSION 1  
KEYWORDS  
SOURCE HTGS\_PHASEL HTGS\_DRAFT HTGS\_FULUTOP.  
ORGANISM human.  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 202993)  
REFERENCE  
1 Birren B., Linton L., Nusbaum C. and Lander E.  
Homo sapiens chromosome 8, clone RP11-678D18  
2 Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,  
Anderson S., Baldwin J., Barina N., Bastien V., Bada P.,  
Bogdanavskiy L., Boukhalter B., Brown A., Burnett G.,  
Campopiano A., Castle A., Choepel Y., Colangelo M., Collins S.,  
Collamore A., Cooke P., DeRellano K., Dewar K., Diaz J.S.,  
Degen S., Domino M., Doyle M., Ferreira P., Fitzhugh W., Gage D.,  
Galegher J., Gardy M., Ginde S., Goyette R., Graham M.,  
Grand-Pierre M., Grant G., Hagos B., Heston A., Horton L.,  
Kane J., Klee J., Johnson R., Jones C., Kamm L., Kerasas A.,  
Klodt J., LaPointe S., Lander E., Lander E., Lander E.,  
Levine R., Lien C., Liu G., Locke K., MacDonald P., Marquis N.,  
McCarthy M., McKean P., McGuire A., McKernan K., McKeeters R.,  
Meldrum J., Menes L., Mihova T., Miranda C., Mlenga V., Morrow J.,  
Murphy T., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,  
O'Neil D., Olyar T.M., Oliver J., Peterson K., Pierre N.,  
Pisani C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D.,  
Roy A., Santos R., Schauer S., Severy P., Spencer B.,  
Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J.,  
Thompson S., Throckmorton J., Travers M., Travers M.,  
Vassiliadis A., Vidal E., Vitell A., Wainwright M.,  
Young G., Zainoun J., Zimmer A. and Zody M.  
Direct Submission  
Submitted (20-APR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 4, 2001 this sequence version replaced g1:7622413.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/rw/RepeatMasker.html  
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: M1BR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center Project name: 750-2  
Center Project ID: 118  
Center Project ID: 118

----- Summary Statistics  
Sequencing vector: M13: M78185: 41% of reads  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap: version 0.960721  
Consensus quality: 202092 bases at least Q40  
Consensus quality: 202411 bases at least Q30  
Consensus quality: 202615 bases at least Q20  
Insert size: 202793; sum-of-ctrls  
Quality coverage: 9.8 in Q20 bases; agtrose-fp  
Quality coverage: 10.1 in Q20.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This is the best possible representation of the data  
\* as soon as it is available and the accession number will  
\* be preserved.

1 6270: contig of 6270 bp in length  
6271 6370: gap of 100 bp  
6371 84296: contig of 77926 bp in length  
84297 84397: gap of 100 bp  
84397 202993: contig of 118597 bp in length.

FEATURES  
Source  
Location/Qualifiers  
1. 192972  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="8"  
/chromosome="8"  
/clone="RP11-678D18"  
/clone.lib="Rpci-11 Human Male BAC"

misc\_feature  
1. 6270  
/note="assembly-fragment  
clone-end:SPect"

misc\_feature  
6371-84296  
/note="assembly-fragment"

misc\_feature  
84397..202993  
/note="assembly-fragment  
clone-end:T7  
vector:side:r19h"

BASE COUNT 58297 a 38729 c 40566 g 65101 t 200 others  
ORIGIN

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-750-590-2 x AC060775 ..

Align seg 1/1 to: AC060775 from: 1 to: 202993

706 LysValCysLeuAspAsnGlyLeuLeuThr 715  
118936 AAGGTCCTTCGTCGACCAACACCTCCTG 118936

seq\_documentation\_block:  
LOCUS CE97611C 313573 bp DNA INV 20-JUN-2001  
DEFINITION Caenorhabditis elegans cosmid Y57G11C, complete sequence.











```

        /db_xref=taxon:10116"
        /tissue_type=livest
        /tissue_tld=lambda gcll1"
    <1..3019
    30..1274
    /codo_start=1
    /product=cytosolic NADP-dependent isocitrate dehydrogenase"
    /protein_id="AA59356.1"
    /db_xref="GI:537941"
    /translation="MSRIKGGSEVEMQGDPMRTIMELIKERILPEVELDLSYDLTRENQTDIDVDVRDAALFKRYGVGKCATITTPBKKRVEEPRLOMKMSSNETIRNPLDGIVPTLNTCTRLTYVGWFRLLTGRIAAGDDIATDVDPNGKVSETIPDKLDGYFVVAVLVVGLISSTLRNDSDLEEDFASSRDMLSKDEPIISTKTLPKRKYEPHDIPOETIDKNXYSPFLSPPFTSRVTDFSLDLYVLDTLVSTLTNRYSODSYSGDSIGLCMTNSVALICPDCKTVTEAQAACAPTHPHMYOGESLTSNPDELINASRCIALHAHLDNLNTELSPFNALLEVCBTIEGFPTFDLAACIXGLPVRSDYLFNEFFMDXKGINSNKATLKAOAK."
    1275..1719
}UTR
ORIGIN      508   a     347 c     413 g     451 t

alignment_scores:
Quality:          9.00           Length:              9
Ratio:            1.000         Percent Identity:    100.000
Percent Similarity: 100.000       Percent Identity:    100.000

Alignment_block:
US-09-750-590-2 x RATCNDI ..

Align seg 1/1 to: RATCNDI from: 1 to: 1719

657 GUAAGAULysAlaLysLeuAlaGIN 665
|||||
1236 GAAGACTTAAGCCCAATTGAAGTCGAG 1262

seq_name=gq_Htg:AC014339

seq_documentation_block:
LOCUS AC014339                2480 bp             DNA                      HTG               16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC014339
VERSION AC014339.1 GI:616596
KEYWORDS HMG, HMGCS_PHRASE.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
EXTRACTORS Metaxera; Athripsoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2480)
AUTHORS Adams M. and Venter J.C.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-1999) Celera Genomics, 45 West Gate Drive, Redwood City, CA 94063, USA identified as CDM:10211476 by the submitter. For further information on this sequence e-mail to fly@celera.com.
COMMENT * NOTE: This is a "working draft" sequence.
* This sequence will be replaced.
* By the finished sequence as soon as it is available and
* The accession number will be preserved.
LocationQualifiers
1..2480
/oranism=Drosophila melanogaster"/db_xref=taxon:"727"
BASE COUNT      748   a     524 c     458 g     750 t
ORIGIN

```

```

Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x AC014339/rev ..

Align seg 1/1 to reverse of: AC014339 from: 1 to: 2480
182 SerArgaspysginsinotricftriala 190
|||||
1979 AGCCGCGATACAAACAGACAGCACT 1953

seq_name: gb.Pr:565762

seg_documentation_block: 3756 bp mRNA PRI 24-NOV-1993
DEFINITION SerArgaspysginsinotricftriala 190
ACCION SS5762
VERSION 565762.1 GI:425552
KEYWORDS
SOURCE ORGANISM
human.
Homio sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
I (bases 1 to 3756)
Chang,J.G., Scarpa,A., Edoy,R.L., Byers,M.G., Harris,A.S.,
Crawford,J.S., Watkins,T.B., Shost,T.B. and Forgez,B.G.
Cloning of the monkeyrhodopsin gene and cDNA for human
beta-todrin, the nonerythroid form of beta-spectrin
Genomics 17 (2), 287-293 (1993)
94010920
Genbank staff at the National Library of Medicine created this
entry [NCBI gblines 13/9/93] from the original journal article.
This sequence comes from Fig. 3.
Map location: 4p21.
Authors state, with three exceptions, this sequence corresponds to
that reported by et al.(1992), Genbank Accession No. M96803.
FEATURES
source
gene
genie
COS
1..3756 /organism=Homo sapiens*
1..3756 /db_xref=taxon:9606*
1..3756 /partial
/gene="SPBNI"
/note="Beta-Tod-In"
1..3756 /partial
/gene="SPBNI"
/note="This sequence comes from Fig. 4"
/product="beta-Todrin"
/protein_id="AAB28324.1"
/db_xref="GI:425553"
/translation="VINDALTEKNIKESLASDLEWTEITILINNKRFSNIVNG
VOODLAFNTFRVEKPEKPTENSLVLFTTOSKMANNRKYVPKRGELSDINRGE
AMBEKLEAKERREINRELINQERLSOLARRDRKARKEETWSEMRGLSDINRGE
VLTTELLEKAREPRLMTGLTSTEDPAVYANDELLEKTLRTLNKNQV RLDA
KITVLEKAEPRERLMELGLTSTEDPAVYANDELLEKTLRTLNKNQV RLDA
KITVLEKAEPRERLMELGLTSTEDPAVYANDELLEKTLRTLNKNQV RLDA
KAERRNLAEERLERKPPMAEEEGEKIEEEKILSSDDGADLYSMILLSHRRAPF
DENSGKGFPOAIKCEGDIAIEHPQSEKIPIRIYLIREOANLDQLSAIRKLREJE
ASLLHDPQADMDIDAMDLDIKTYSVDVGHQRYSTOSLVSKKVDAEELIAMPPTLI
DETHLOASHLOJSHAEPDKRGRLSGIEREKVEVALTRLRQAOLDPLIALYMFSEA
DACBELADEKQOANNMQ IPKKULDEVIHQRESLEPENNNASRAVAVNQIARQL
HSGPSKEIPKAOQRRLTWMSOFRELYDRKALLASTTOYHLCEKTKSNIRK
HSESEEDONKNSGNWALQKRLRDEVDVAIEKLSUDREKLESSEPHQNQQ
TISSEEDONKNSGNWALQKRLRDEVDVAIEKLSUDREKLESSEPHQNQQ
PTITTAPEKILTOHKENTIKENTIDYEDONKDEKENVTOGCDONWEPTLHQLSL
TOMNELKHMKNEMLNQSOSHUYOFLDPTOKAEFLANOEXVLAHIEPTTLEKAEK
AARKOEOPDTTHADANEK IAAVETGRVLSDENINSRLOKDYSIDORRKRPFAAL
SELMLAKNKNDIDLODCDELALVINKEKLLADNSDYOEARNLSIKWLHQAPPAAL
LASKRWKLL IEKEMQDI LSEKPETAVVVKETKLTGHMEVLESTTGOTPDLPADA
KAEFLTSCADSLQADILEGQISDGYKQUTSVNILLKQOALKENQKREKETEBE
```





```

alignment_scores:
  Quality: 9.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Percent Identity: 100.000

alignment_block:
  US-09-750-590-2 x HUMSPTB1A ..
  Align seg 1/1 to: HUMSPTB1A from: 1 to: 7561
  696 LysAndIothrLeuGlnIlySGluIle 704
  4886 AAAATCGACCCCTCCACAAAGAAATC 4912

```

---





Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-750-590-2 x AAH40197 ..

Align seg 1/1 to: AAH40197 from: 1 to: 1980

```

47 AAMPpGLYSLAASPAVIGLGLAYGSESLAPHEHSLAYALAI 63
|||||
184 AATCCAGCAAACTAGATGTGAGAGCAATCTCCCTTCACATTTGTGCG 233
|||||
63 ASETLYAGLYAALSLGSLAASLALALSLALSLGSLAYALA 80
|||||
234 CTCAAGGGCAATCTTGAAATGTTGAATCCATCTTATCAATGCACTTG 283
|||||
80 SPLELTHSERAPRTHALAGLYATGASALALSLALSLALAI 96
|||||
284 AATTACAAACGACGACACGACGAGAAATGCTCTTCACTTGCTGCGA 333
|||||
97 LYSTYGLYHSLAALSLGSLAALSLGSLAALSLGSLAYALA 113
|||||
334 AAGTATGGCATGATCTGTCTACAAAACCTCTACAGTCAATGTCTCC 383
|||||
113 OFRGLQHLIA 116
|||||
384 CACTGACAT 393
|||||

```

seq\_name: /SID52/gcgcdata/geneseq/geneseq/NAI2001.DAT:AAH75522

seq\_documentation\_block:  
ID AAH75522 standard: cDNA: 2222 BP.

AAH75522:

18-OCT-2001 (first entry)

Human myosin heavy chain 74 encoding cDNA.

Human: myosin heavy chain 74; disease: cancer; HIV; infection;

human immunodeficiency virus; ss.

Homo sapiens.

Location/Qualifiers

key CDS /tag= a /product= "myosin heavy chain 74"

note= "claimed in claim 6"

CN1296972-A.

30-MAY-2001.

23-NOV-1999: 99CN-0124081.

23-NOV-1999: 99CN-0124081.

(SHAN-) SHANGHAI BOHONG GENE DEV CO LTD.

MOO Y, Xie Y;

WPI: 2001-483899/53.

P-PSDB: AAG64560.

Polypeptide-human myosin heavy chain 74 and polynucleotide for coding

said polypeptide -

Claim 6; Page 25-26 Disclosure: 36pp: Chinese.

CC cDNA.

Sequence 2222 BP: 905 A: 351 C: 479 G: 487 T: 0 other:

alignment\_scores:

Quality: 53.00 Length: 53  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-750-590-2 x AAH75522 ..

Align seg 1/1 to: AAH75522 from: 1 to: 2222

```

1274 GLUATGYSAPLYSSELEUTHETHTHLETHGLULSLAALGARGII 1290
|||||
1273 GAACAGTGTGATAGCTCTTACACACACACACACACACACACACAC 1772
|||||
1290 EGLGLUSERLALYSGHILSLGLALALYASAPASLYSLLETHGLUL 1307
|||||
1773 ACAGAACTGCTTAAACAAATATGAAACCAAAAGATTAATGATTAAGT 1822
|||||
1307 EULASAPASVAGLUATGSLAULSLGSLALALSLAASNGLYLEUSGIN 1323
|||||
1823 TCCCTTAATGATGTGAAAGATTTAAACACGACACTCAATGCTTCCCA 1872
|||||
1324 LEUTHTYT 1326
|||||
1873 CTCACCTTAC 1881
|||||

```

seq\_name: /SID52/gcgcdata/geneseq/geneseq/NAI1999.DAT:AAV83134

seq\_documentation\_block:  
ID AAV83134 standard: cDNA: 2322 BP.

AAV83134:

02-MAR-1999 (first entry)

Polynucleotide clone du410\_5 encoding a secreted protein.

Clone: secreted protein; protein factors; cytokine; lymphokine;

Interferons; colony stimulating factor; CSF; interleukin; cloning;

tumour invasion; tumour suppression; immune boosting; ds.

Homo sapiens.

key CDS /tag= a /product= Secreted\_protein

W09843302-A1.

05-NOV-1998.

24-APR-1998: 98MO-US08336.

23-APR-1998: 98US-0065125.

25-APR-1997: 97US-0845296.

(GENM ) GENETICS INST INC.

Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

Recie LA, Spaulding V, Treacy M;

WPI: 1999-024059/02.

P-PSDB: AAV85608.

New polynucleotides encoding secreted human proteins - are derived from human foetal brain, adult brain, adult blood or placenta cDNA libraries, useful, e.g. as potential immunomodulators



```

PE      30-JUN-2000 : 2000WU-USI8374 .
PR      02-JUL-1999 :   99US-0142310 .
PR      02-JUL-1999 :   99US-0142311 .
XX      PA
PA      (CHIR ) CHIRON CORP .
PA      (HYSE-) HYSED INC .
PI      Williams LV , Escobedo J , Ionis MA , Garcia PD , Klingler J , Kassam A ;
XN      Reinhard C , Randazzo F , Kennedy GC , Pot D , Lamson G , Dimanic R ;
PI      Chkentalov K , Dimanic S , Dickson M , Labat I , Leshkowicz D ;
XN      Kleit D , Garcia V , Jones LW , Strache-Crain B ;
PI      WPI : 2001-091805/10 .
DR
XX      Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer specifically of the colon or
PT prostate, comprises 3531 human polynucleotide sequences -
XX
PS      Claim 9: Page 805; 1046pp: English .
XX
CC      The present sequence is one of 3351 sequences in a library of human
CC polynucleotides . The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer . The library can be used
CC for detection of expression levels of mRNA and so produce additional copies
CC of the polynucleotides for detection of transcription levels . Ribosomes
CC or antisense oligonucleotides can be generated . The polynucleotides and/or
CC their gene products are used as genetic or biochemical markers (e.g. In
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions . The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasias .
XX
SO      Sequence 401 BP: 114 A; 85 G; 107 G; 93 T; 2 other:

alignment_scores:
    Quality:       47.00     Length:        47
                   Ratio:    1.000           Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x AMF66074 ..

Align seg 1/1 to: AMF66074 From: 1 to: 401

118 AsPLauDlncGlyAghfAlalAuLluHlaAsPaLaAlwMeLAAsPCyPr 134
..
45 GAACCTGCAGCGGAACAACGTCACTTCATGACTGGCAGAAAGCCAGAATTGCC 94
134 oGeSerTtTeGLllleLuLnCyDeAtgRhlISdYpALasrVtAlsmnLLMyra 151
..
95 TTCTAGCATMACAGTGCTCTTTTGTAOAcATCGGCGCTCTGTCTANTGCGAAAG 144

151 sPyvAIAPCIYAqgrTHrrFLreLvWvlJueuAlArTrCLImcE 164
145 ATGTAGACGGCGGAGAACACCATTGTGCTGGCTMACTCAATG 185

seq_name:= /SID$2/gcgdata/geneseq/geneseqn/NA2001.DAT: NA129101
seg_documentation_block:
XX      NA129101 standard, cDNA; 441 bp.
XX      AA129101:
DT      12-OCT-2001 (first entry)
XX
XX      Colon tumour related determined cDNA sequence for clone CT2-55.
```

KW Human: Immunotherapy; diagnosis; colon cancer; colon tumour;  
 KM immunogenic; gene therapy; vaccine; colonic Cancer; ss.  
 XX Homo sapiens.  
 XX MO200149716-A2.  
 XX  
 PD 12-Jul-2001.  
 PF 29-Dec-2000; 2000MO-US55596.  
 PR 30-Dec-1999; 99US-0476296.  
 PR 10-Jan-2000; 2000US-0480321.  
 PR 15-Feb-2000; 2000US-0504529.  
 PR 06-Mar-2000; 2000US-0519444.  
 PR 26-Mar-2000; 2000US-0519444.  
 PR 26-Jun-2000; 2000US-0659448.  
 PR 28-Aug-2000; 2000US-0649811.  
 PA (CORI ) CORIXA CORP.  
 PX Xu J., Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jiang Y;  
 DR WPI: 2001-441847/47.  
 XX Colon tumor associated proteins and nucleic acids useful for the  
 PT prevention, diagnosis and treatment of colonic cancer .  
 PS Claim 25; Page 308; 47pp: English.  
 XX The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity,  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression. (I) and (II) may be used in the prevention, diagnosis and  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC patients own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC identify TCAP genes. (I) and (II) may be used in the prevention and  
 CC therefore which patients may be in need of restorative therapy (I) may  
 CC also be used as identifiers in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512  
 CC given in the exemplification of the present invention.  
 XX Sequence 441 BP: 127 A; 97 C; 114 G; 102 T; 1 other:  
 SO  
 Alignment\_scores:  
 Quality: 47.00 Length: 47  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 Alignment\_block:  
 US-09-750-590-2 x AA129101 ..  
 Align seg 1/1 to: AA129101 from: 1 to: 441  
 118 AsplENGInLYArghThrlAlcLmHlSAsPAAlaAlaMeLAaapCysPr 134  
 |||||||.....  
 137 GAQCTGCGAGGAACACTGCATTCACGATGCCGCAATGGCATGTGTGCC 186

134 oserSerIleGlnLeuLeuCYAsPhtAsIcLYAlAserValAsnAlaLYSA 151  
 |||||||  
 187 TTCAGCATACACCTGCTTGTGGACCATGGGGCTCTGTAAATCCAAAG 236  
 |||||||  
 151 sPVAlaSPcLYaGThThPrIealValLeuAlaThGlnMet 164  
 |||||||  
 237 ATGTACAGCGGCGGACACACCTGCTGTGGCTACTCAAGATG 277  
 seq.name: /SID52/5cgdata/geneseq/geneseq/NM2000.DAT:AAA5014  
 seq.documentation.block:  
 ID AAA5014 standard: cDNA: 571 BP.  
 AC AAA5014;  
 XX  
 XX 21-AUG-2000 (first entry)  
 DT  
 XX  
 XX Human secreted expressed sequence tag SEQ ID NO:1569.  
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;  
 KW expressed sequence tag; EST; probe; chemokine; proliferative;  
 KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;  
 KW thrombolytic; antiinflammatory; cytotoxic; antibacterial; antifungal;  
 KW antiviral; antidiabetic; antisthmatic; vulnere; antiparkinsonian;  
 KW antitumor; osteopathic; neuroprotective; noctropic; antipsoritic;  
 KW ceruloprotective; anticoagulant; antidepressant; gene therapy;  
 KW vaccine; autoimmune disorder; multiple solid cell deficiency; ulcer;  
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KW central nervous system disorder; Alzheimer's disease; stroke;  
 KW Parkinson's disease; Huntington's disease; coagulation disorder;  
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KW tumour; infection; depression; psoriasis; ss.  
 KW  
 KW Homo sapiens.  
 OS  
 XX  
 XX NO20002191.X1.  
 XX  
 XX 20-APR-2000.  
 PD  
 XX  
 XX 15-OCT-1999; 99MO-US24206.  
 PF  
 XX  
 XX 15-OCT-1998; 98US-0104436.  
 PR  
 XX  
 XX (GENY ) GENETICS INST INC.  
 PA  
 XX  
 XX Jacobo F. McCoy JM, Lavalley ER, Collins-Racle LA, Evans C;  
 PI Metberg D, Treacy M, Bowman MR;  
 PT  
 XX  
 XX MPI: 2000-317938/27.  
 DR  
 XX  
 XX Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (eSSTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 PT  
 XX  
 XX Claim 1: Page 616, 803pp: English.  
 XX  
 XX AAA3426 to AAA4592s represent specifically claimed secreted expressed  
 CC sequence tags (eSSTs). Isolated from human, mouse, chicken and rat  
 CC tissue sources. The eSSTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC Chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC osteoclastic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytotoxic; antidiabetic; antisthmatic; vulnere; antiparkinsonian;  
 CC antitumor; osteopathic; neuroprotective; noctropic; antipsoritic;  
 CC ceruloprotective; anticoagulant; antidepressant; gene therapy;  
 CC vaccine; autoimmune disorder; multiple solid cell deficiency; ulcer;  
 CC lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 CC central nervous system disorder; Alzheimer's disease; stroke;  
 CC Parkinson's disease; Huntington's disease; coagulation disorder;  
 CC haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 CC tumour; infection; depression; psoriasis; ss.  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions

CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and  
 CC prolasias. AAA4526 to AAA4592s represent variants which are given  
 CC in the exemplification of the present invention.  
 XX  
 XX Sequence 571 BP: 246 A: 80 C: 125 G: 120 T: 0 other:  
 SQ  
 Alignment scores:  
 quality: 41.00 Length: 41  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 Alignment block:  
 US-09-750-590-2 x AAA5014 ..  
 Align seq 1/1 to: AAA5014 from: 1 to: 571  
 1274 GUAATCYAsAPuYSerIeThThPrIeThGlnLeuGlnArgI1 1290  
 |||||||  
 446 GACGATGTGTAATGTCCTTACCAACATCAACGATTCACCAAGAAAGAT 485  
 |||||||  
 1290 eGlnGSerIaLYsGlnIleGlnAlaLYsAPuYSerIeThG1u1 1307  
 |||||||  
 496 ACAGATCTGCTTAAACAAATGAAACCAAAATATATACATTAAGTGAAC 545  
 |||||||  
 1307 eULeAsnAPuYAlGlnArgLeu 1314  
 |||||||  
 546 TCGTTAATGATGTGGAAGGACTC 568  
 seq.name: /SID52/5cgdata/geneseq/geneseq/NM2000.DAT:AAA77915  
 seq.documentation.block:  
 ID AAA77915 standard: cDNA: 421 BP.  
 AC AAA77915;  
 XX  
 XX 14-NOV-2000 (first entry)  
 DT  
 XX  
 XX cDNA encoding human colon tumour protein CT607, SEQ ID NO:195.  
 DE  
 XX  
 XX Human colon tumour polypeptide; tumour antigen; cancer; vaccine;  
 KW immunotherapy; diagnosis; progression; ss.  
 KW  
 KW Homo sapiens.  
 OS  
 XX  
 XX NO200037643-A2.  
 PN  
 XX  
 XX 29-JUN-2000.  
 PD  
 XX  
 XX 23-DEC-1999; 99MO-US30909.  
 PR  
 XX  
 XX 23-DEC-1998; 98US-0221298.  
 PR  
 XX  
 XX 02-JUL-1999; 99US-0347496.  
 PR  
 XX  
 XX 22-SEP-1999; 99US-0401064.  
 PR  
 XX  
 XX 19-NOV-1999; 99US-0444242.  
 PR  
 XX  
 XX 02-DEC-1999; 99US-0454150.  
 PR  
 XX  
 XX (COR1-) CORIXA CORP.  
 PA  
 XX  
 XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolt J;  
 PI Wang T, Yugu J;  
 PT  
 XX  
 XX MPI: 2000-442671/38.  
 DR  
 XX  
 XX New colon tumor polypeptides used to inhibit the development of cancer,  
 XX especially colon cancer, and for diagnosing and monitoring the  
 PT progression of the cancer.  
 PT  
 XX



ID AI129076 standard; cDNA: 421 BP.  
 AC AI129076;  
 XX  
 DT 12-OCT-2001 (first entry)  
 DE Colon tumour related determined cDNA sequence for clone CT607.  
 XX  
 XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 XX immunogenic; gene therapy; vaccine; colonic cancer; ss.  
 XX Homo sapiens.  
 XX MO200149716-A2.  
 XX  
 XX 12-JUL-2001.  
 PD 29-DEC-2000; 2000MC-US35596.  
 PF 30-DEC-1999; 99US-0476286.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 XX  
 XX (COR1-) CORTXA CORP.  
 XX  
 XX Xu J, Lodes MJ, Secrist H, Benson DR, Neagher MJ, Stolk JA,  
 PI Kling GE, Wang Y;  
 XX WPI: 2001-441847/47.  
 DR  
 XX Colon tumor associated proteins and nucleic acids useful for the  
 XX prevention, diagnosis and treatment of colonic cancer.  
 XX  
 XX Claim 25; Page 300; 47pp; English.  
 CC The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (I) and (II) may be  
 CC used in the prevention, diagnosis and treatment of diseases associated  
 CC with inappropriate colon tumour associated protein (TCAP) expression.  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC patients own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC and therapeutic applications. The present invention also includes a method  
 CC to mutate the presence of similar nucleic acids in samples, and (I) may  
 CC also be used as antigens in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AI128460 to AI129512  
 CC and AA04494 to AA04523 represent nucleotide and amino acid sequences  
 CC given in the exemplification of the present invention.  
 XX  
 XX Sequence 421 BP: 109 A: 101 C: 125 G: 86 T: 0 other;  
 alignment\_scores:  
 Quality: 37.00 Length: 37  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:

US-09-750-590-2 x AI129076 ..  
 Align seg 1/1 to: AI129076 from: 1 to: 421  
 64 SerTSGIYanIeugIucYsIeuaAaA1a1leu1e1h1e1s1g1y1a1s 80  
 310 TCAAGGGGAATCTTGAGCTTTGAATCCCACTCTATACATGGAAGCTTGA 359  
 80 P11e1h1e1S1e1S1e1P1h1A1a1G1Y1a1e1u1e1S1e1S1e1S1e1A1a1 97  
 360 TATTACACCGATGACCTGCGAGGAGAAATGCTCTTCAACGCTGCTGCTA 409  
 97 ySTyG1yH1s 100  
 410 AGTATGACAT 420  
 seq\_name: /SDS2/gcdata/geneseq/geneseqn/NA2000.DMT-AAC27789  
 seg\_documentation\_block:  
 ID AAC27789 standard; cDNA: 361 BP.  
 AC AAC27789;  
 XX  
 DT 06-OCT-2000 (first entry)  
 DE Human secreted protein 5' EST, SEQ ID NO: 31864.  
 XX  
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 XX gene therapy; chromosome mapping; ss.  
 XX Homo sapiens.  
 XX  
 XX EP1033401-A2.  
 PD 06-SEP-2000.  
 PR 21-FEB-2000; 2000EP-0200610.  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 XX (GSEST) GENSET.  
 XX  
 XX Dumas Milne Edwards J, Duclet A, Giordano J;  
 XX WPI: 2000-500381/45.  
 DR  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 XX Claim 1: SEQ ID 31864; 71pp + CD-ROM; English.  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC human cells. The 5' ESTs have not been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the 5' ends of the cDNA are therefore  
 CC usually missing. The full length cDNAs and genomic DNAs 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 XX Sequence 361 BP: 112 A: 75 C: 89 G: 83 T: 2 other;  
 alignment\_scores:  
 Quality: 23.00 Length: 23  
 Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

us-09-750-590-2 x AAC27789 ..

Align seq 1/1 to: AAC27789 from: 1 to: 361

118 AspleuglmglyAqGthFAlaLeuHlAmpAlaAlaMeAlaAspCyPr 134  
 |||  
 54 GACCTGCAGCGAAGACGACACTACAGATGCCCAATGCGACATTCGCC 103  
 |||  
 134 GGGGAGTlleglInleuAla 140  
 |||  
 104 TTTCACGTCACGTCGCTT 122

seq\_name: /SID52/gcdata/geneseq/geneseqn/NA2000.DAT:AAC27957

seq\_documentation\_block:

ID AAC27957 standard; cDNA: 355 BP.

AC AAC27957:

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 32032.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN BP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

PR (GEST ) GENSET.

PI Dumas MLine Edwards J, Duclet A, Giordano J;

DR WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX claim 1: SEQ ID 32032; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC cDNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC that encode the 3' ends of mRNAs and even in those cases where longer  
 CC cDNAs have been obtained, the full 5' ends of the 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNA with the full 5' ends. 5' ESTs are  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are  
 CC used in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

Sequence 355 BP; 110 A; 61 C; 62 G; 120 T; 2 other;

alignment\_scores:

Quality: 22.00 Length: 22  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

us-09-750-590-2 x AAC27957 ..

Align seq 1/1 to: AAC27957 from: 1 to: 355

1355 AsplaAspAqgGlnHsglIngluValIleAlaIleIryrA-ThrHsle 1371  
 |||  
 122 GATCTGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 171  
 |||  
 1371 uleuSerAlaAlaIaIn 1376  
 |||  
 172 TCTTAGCTGCTGCACAG 187

seq\_name: /SID52/gcdata/geneseq/geneseqn/NA1999.DAT:AA41086

seq\_documentation\_block:

ID AA41086 standard; cDNA: 356 BP.

AC AA41086:

XX 18-JUN-1999 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 298.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW for therapy; chromosome mapping; signal peptide; secretion;  
 KW upstream regulatory sequence; haematopoiesis regulation; tissue growth regulation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.

PN HQ9906554-A2.

XX 11-FEB-1999.

PF 31-JUL-1998; 98MO-1B01238.

XX 01-AUG-1997; 97US-0905134.

PR (GEST ) GENSET.

PI Duclet A, Dumas MLine Edwards J, Lacroix B;

DR WPI: 1999-153784/13.

DR P-PSDB: AA412253.

XX New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from kidney, fetal kidney, dystrophic  
 PT muscle, muscle and heart tissue  
 XX claim 1: Page 442; 622pp; English.

XX AA410826 to AA411093 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AA41042 and  
 CC AA411994 to AA412260, respectively. The proteins given represent the  
 CC signal peptide and an N-terminal fragment of a secreted protein. The  
 CC nucleic acid sequences can be used for producing secreted human gene  
 CC products. They can also be used to develop cytokine activity, cell  
 CC therapy. The proteins obtained may have cytokine activity, cell  
 CC therapy, secreted protein activity, haematopoiesis regulating  
 CC activity, tissue growth regulation, haematopoiesis regulating  
 CC activity, thrombolytic/chemotherapeutic activity, haematopoietic  
 CC regulating activity, chemotherapeutic/chemotherapeutic activity, haematopoietic  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used  
 CC for directing extracellular secretion of the polypeptide or the insertion  
 CC of polypeptide into a membrane, or importing a polypeptide into  
 CC a cell.









```

seq_name: /SID2/sgcdata/geneseq/geneseq/MA2001.DAT:AAH33568
seq_documentation_block:
ID AAH33568 standard; cDNA: 653 BP.
AC AAH33568;
XX
XX
XX 03-SEP-2001 (first entry)
XX
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:624.
DE Human colon cancer antigen; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
XX MO200123292-42.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000: 2000MO-US26524.
XX
XX 29-SEP-1999: 990US-0157137.
XX
XX 03-NOV-1999: 990US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Blase CE, Rosen CA;
PI WPI: 2001-235357/24.
XX
XX P-PSDB: AAC74137.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1: Page 2661: 9803P: English.
XX
XX AAH32943 to AAH37195 and AAC73514 to AAC77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy to combat colorectal diseases and may be used in the prevention,
XX diagnosis and treatment of colorectal diseases. The nucleic acid P
XX expression. For example, N and P may be used to treat colorectal P
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing P.
XX Additionally, N may be used to produce the colon cancer-associated Pa,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins N and P. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal diseases. AAH37196 to AAH37204
XX and AAC7789 represent sequences used in the exemplification of the
XX present invention.
XX
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 653 BP: 216 A; 84 C; 103 G; 243 T; 7 other:
XX
XX
XX alignment_scores:
XX Quality: 13.00 Length: 13
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-750-550-2 x AAH33568
XX
XX Align seg 1/1 to: AAH33568 from: 1 to: 653
XX
XX 376 LysGLuSPMeLLeuLLeuLysGLInGInMeTyrMet 388
XX |||||||
XX 571 AAAGAGATATCTCTCTTAACAGCGTCAATGATATAG 609

```

```

seq_name: /SID2/sgcdata/geneseq/geneseq/MA2001.DAT:AA159207
seq_documentation_block:
ID AA159207 standard; cDNA: 1712 BP.
XX
XX AA159207;
XX
XX 22-OCT-2001 (first entry)
XX
XX
XX Human polynucleotide SEQ ID NO 1410.
XX
XX Human: neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX MO200153112-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000: 2000MO-US34263.
XX
XX 21-JAN-2000: 2000US-0488725.
XX
XX 25-APR-2000: 2000US-0552147.
XX
XX 09-JUL-2000: 2000US-0598042.
XX
XX 19-JUL-2000: 2000US-0620312.
XX
XX 03-AUG-2000: 2000US-0653450.
XX
XX 14-SEP-2000: 2000US-0662191.
XX
XX 18-OCT-2000: 2000US-0693036.
XX
XX 29-NOV-2000: 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Aunnd V, Chen R, Mo Y, Qian XB, Ren F, Wang D;
XX Wang J, Mehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QN, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX
XX P-PSDB: AAH40051.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1: SEQ ID NO 1410: 10078BP: English.
XX
XX
XX The invention relates to human nucleic acids (AA15798-AA161369) and
XX the encoded polypeptides (AAH30642-AAH42213) with neurotrophic
XX activity. The polynucleotides are useful
XX in gene therapy. A composition containing the polynucleotides
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX modulation of activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, chemotactic/chemokinetic activity, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemia and
XX C.N.S disorders.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 1712 BP: 367 A; 521 C; 555 G; 269 T; 0 other:
XX
XX
XX alignment_scores:
XX Quality: 11.00 Length: 11
XX Ratio: 1.000 Gaps: 0

```

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x AAI59207 ..

Align seq 1/1 to: AAI59207 from: 1 to: 1712

90 AenAlaLeuHisIeuAlaAlaLysTyrGlyHis 100  
|||||  
439 AATGCCCTCCACCTGGCCGCAAAATAGCGCAC 471

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:AAQC94930

seq\_documentation\_block:

ID AAC94930 standard: cDNA; 549 BP.

AAQC94930:

19-FEB-2001 (first entry)

Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:1425.

Cat flea: hindgut and Malpighian tubule nucleic acid; HMT;

flea infestation: vaccine; antiparasitic; therapeutic target;

diagnosis: detection; ss.

Ctenocephalides felis.

MO200001621-A2.

19-OCT-2000.

07-APR-2000: 2000MO-US09437.

09-APR-1999: 9905-0128704.

(HESKA) HESKA CORP.

Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;

WPI: 2000-656323/63.

Flea Malpighian tubule and head and nerve cord tissue derived nucleic

acid sequences -

Intestations -

Intestations -

Claim 26: Page 709-710; 964pp: English.

The invention relates to novel cat flea (Ctenocephalides felis) nucleic

acids which are expressed in hindgut and Malpighian tubule (HMT) tissue

or head and nerve cord (HNC) tissue. The invention also relates to the

encoded proteins. The invention additionally encompasses expression

constructs, recombinant viruses and recombinant cells comprising the

nucleic acids of the invention, recombinant production of the proteins,

antibodies against the proteins, a method of identifying inhibitors of

the proteins, a method of identifying inhibitors of the proteins, a method

of identifying inhibitors of the proteins, a method of identifying inhibitors

of the proteins, a method of identifying inhibitors of the proteins, a method

of identifying inhibitors of the proteins, a method of identifying inhibitors

of the proteins, a method of identifying inhibitors of the proteins, a method

of identifying inhibitors of the proteins, a method of identifying inhibitors

CC present sequence represents a cat flea HMT cDNA of the invention.

XX

XX

Sequence 549 BP: 194 A; 80 C; 114 G; 159 T; 2 other;

SO

alignment\_scores:

Quality: 9.00 Length: 9

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x AAC94930 ..

Align seq 1/1 to: AAC94930 from: 1 to: 549

1020 HleGluThrLeuGlnLysAspLeuLys 1028

18 AATATACCTACACAAAGACTTAAG 44

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ20077

seq\_documentation\_block:

ID AAZ20077 standard: DNA; 715 BP.

AAZ20077:

20-APR-1999 (first entry)

Enterococcus faecalis EF040 gene fragment.

Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

detection; attenuation; antigenic; ss.

Enterococcus faecalis.

MO9850554-A2.

12-NOV-1998.

04-NOV-1998: 98MO-US08959.

14-NOV-1997: 97US-0065009.

06-MAY-1997: 97US-0044031.

16-MAY-1997: 97US-0046555.

(HUMA-) HUMAN GENOME SCI INC.

Bailey C, Choi GH, Hromockyj A, Kunsch CA;

WPI: 1999-070095/06.

P-PSDB: AAY00087.

New isolated Enterococcus faecalis polynucleotides - used to develop

products for the detection and diagnosis of Enterococcus infection

for prevention or attenuation of Enterococcus infection

Claim 1: Page 119; 301pp: English.

The present sequence encodes an antigenic polypeptide fragment

isolated from Enterococcus faecalis. The present invention describes

genes, proteins and antigenic polypeptides isolated from faecal

alignment\_scores:

Quality: 9.00 Length: 9

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x AAC94930 ..

Align seq 1/1 to: AAC94930 from: 1 to: 549

1020 HleGluThrLeuGlnLysAspLeuLys 1028

18 AATATACCTACACAAAGACTTAAG 44

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ20077

seq\_documentation\_block:

ID AAZ20077 standard: DNA; 715 BP.

AAZ20077:

20-APR-1999 (first entry)

Enterococcus faecalis EF040 gene fragment.

Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

detection; attenuation; antigenic; ss.

Enterococcus faecalis.

MO9850554-A2.

12-NOV-1998.

04-NOV-1998: 98MO-US08959.

14-NOV-1997: 97US-0065009.

06-MAY-1997: 97US-0044031.

16-MAY-1997: 97US-0046555.

(HUMA-) HUMAN GENOME SCI INC.

Bailey C, Choi GH, Hromockyj A, Kunsch CA;

WPI: 1999-070095/06.

P-PSDB: AAY00087.

New isolated Enterococcus faecalis polynucleotides - used to develop

products for the detection and diagnosis of Enterococcus infection

for prevention or attenuation of Enterococcus infection

Claim 1: Page 119; 301pp: English.

The present sequence encodes an antigenic polypeptide fragment

isolated from Enterococcus faecalis. The present invention describes

genes, proteins and antigenic polypeptides isolated from faecal

alignment\_scores:

Quality: 9.00 Length: 9

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x AAC94930 ..

```

Quality:          9.00      Length:          0
Ratio:            1.000     Caps:             9
Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
US-09-750-590-2 x AAX20077    ..

Align seg 1/1 to: AAX20077 from: 1 to: 715

      173 Leu1leAspaRtGlyAlaAspIleAan 181
      |||||
      284 TTGATTGATCCGGCGGCCGCAATTTAAAT 310

seq_name: /SID52/gcgdate/geneseq/genaseqn/MN1999.DMT.AAX20076
seq_documentation_block:
ID ID AAX20076 standard; DNA; 864 BP.
XX AC AAX20076;
XX AC
XX AC 20-APR-1999 (first entry)
XX DT
XX DT
XX DD Enterococcus faecalis gene EF040.
XX KM Enterococcus faecalis infection; vaccine; immune response; diagnosis
XX KM detection; attenuation; antigenic; SS.
XX XX
XX OS Enterococcus faecalis.
XX PN
XX PN MO9850554-A2.
XX PD
XX PD 12-NOV-1998.
XX PE
XX PE 04-MAY-1998; 98NO-US08959.
XX PR
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046555.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PP
XX PP Bailey C, Choi GH, Hromocky J A, Kunsch CA;
DR WPI: 1999-070093/06.
DR P-PDB: AAU00086.
XX PT New isolated Enterococcus faecalis polynucleotides - used to develop
XX PT products for the detection of Enterococcus and for use in vaccines
XX PT for prevention or attenuation of Enterococcus infection
XX PS
XX PS Claim 1: Page 118-119; 301pp; English.
XX CC The present sequence represents a gene isolated from
XX CC Enterococcus faecalis. The protein encoded by this gene has been shown to
XX CC and antigenic polypeptides isolated from this gene are useful as
XX CC be used in vaccines for preventing or attenuating infections caused
XX CC by a member of the Enterococcus genus in an animal. They can also be
XX CC used for detecting Enterococcus antibodies in a sample. The nucleotide
XX CC sequences can be used for detecting Enterococcus nucleic acids.
XX CC Products from the present invention can also be used for screening
XX CC compounds to identify agonists and antagonists of E. faecalis protein
XX CC activity.
XX SK
XX SK Sequence 864 BP; 313 A; 153 C; 195 G; 203 T; 0 other;

Alignment_scores:
Quality:          9.00      length:           9
Ratio:            1.000     caps:              0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x AAX20076    ..

```

```

Align seg 1/1 to: AAM20076 from: 1 to: 864
173 LeuTleasPnTGClYALAAASlileAn 181
|||||
|||
430 TTTGTTGGATCGCAGCGCATATTAT 456
seq_name: /SID82/gcgdata/geneseq/geneseq/AAL999.DAT:AAM64424
seq_documentation_block:
ID AAM64424 standard; cDNA: 3377 BP.
XX AAM64424:
DT 01-MAR-1999 (filed)
DE Mouse elf cDNA.
XX
XX Elf: mouse; liver development; signal transduction; biliary tree;
KM cholestasis; biliary stone; hepatic obstruction; stricture;
KM primary biliary cirrhosis; primary sclerosing cholangitis; therapy;
KM 85.
OS Mus sp.
XX
XX Key Location/Qualifiers
FH CDS 333..6797
FT /tag= a
FT misc_difference 794..1080
FT /tag= /note=
FT misc_difference 2376..3377
FT /tag= b
FT /note=
FT misc_difference 2376..3377
PA W09848827-AI.
XX
PD 05-NOV-1998.
XX
XX 30-APR-1998: 98NC-U508656.
XX
XX 30-APR-1997: 97US-0041349.
PA (MISH/) MISHRA L.
PI Mishra L.;
DR WP1: 1999-009382/01.
XX P-PDOB: AAM61642.
PT
PT New isolated early liver development gene - used to develop
PT products for treating e.g. liver diseases hepatocellular carcinoma
PT degenerative neurological disorders, anemias, ataxia or
PT haemochromatosis
PS
PS Example 2: Fig 16a: 32pp: English.
CC This partial nucleic acid sequence (lacking the 5' end) encodes an
CC elf protein (see AAM61642) of the foetal murine liver. The invention
CC provides early developing liver proteins (see AAM61658-42) and the
CC embryonic liver cDNA library (see AAM64410-24). To isolate the genes, 1, 5,
CC 12.5 and 14.5 post-coital embryos were screened with probes derived
CC from growth factors (IGF-I, IGF-II and IGFBP-2) and transcriptional
CC activators (c/BMP and LRF1), known to be expressed in the developing
CC liver. Subtractive hybridisation was subsequently performed, and
CC isolated clones were analysed for stage specificity by Southern

```

|      |                                                                          |
|------|--------------------------------------------------------------------------|
| CC   | known to be expressed in the developing liver. Subtractive               |
| CC   | hybridisation was subsequently performed, and isolated clones were       |
| CC   | analysed for stage specificity by Southern blotting. 361 genes           |
| CC   | transcript size, abundance and tissue specificity were determined        |
| CC   | (see AAV6411). The cDNA for the c1f protein (see AAV61539-40)            |
| CC   | was found to be important for the formation of the biliary tree during   |
| CC   | early liver development. They are used in a claimed method to            |
| CC   | treat cholestasis, biliary stones, liver obstruction, stricture,         |
| CC   | primary biliary cirrhosis and primary sclerosing cholangitis.            |
| CC   | Early developing liver proteins and nucleic acids can also be used       |
| CC   | in the diagnosis of liver diseases and other disorders, including        |
| CC   | those relating to oncogenesis and tissue repair.                         |
| CC   |                                                                          |
| XX   | Sequence 6991 BP: 2022 A: 1601 C: 1913 G: 1455 T: 0 other:               |
| XX   |                                                                          |
| XX   | alignment_scores:                                                        |
| XX   | Quality: 9.00 Length: 9                                                  |
| XX   | Ratio: 1.000 Gaps: 0                                                     |
| XX   | Percent Similarity: 100.000 Percent Identity: 100.000                    |
| XX   |                                                                          |
| XX   | alignment_block:                                                         |
| XX   | US-09-750-590-2 x AAV64412 ..                                            |
| XX   |                                                                          |
| XX   | Align seg 1/1 to: AAV64412 from: 1 to: 6991                              |
| XX   |                                                                          |
| XX   | 696 LysAGCAGphrGAGGlyGlyGlu 704                                          |
| XX   |                                                                          |
| XX   | 4869 AAAACCGAGCCCTCCGAAAGAAATC 4895                                      |
| XX   |                                                                          |
| XX   | seq_name: /SID52/gcgdata4/geneseq/geneseq/NA2001.DAT:AA504273            |
| XX   |                                                                          |
| XX   | seq_documentation_block:                                                 |
| ID   | AA504273 standard; cDNA; 7240 BP.                                        |
| XX   |                                                                          |
| XX   | AA504273:                                                                |
| DT   |                                                                          |
| XX   | 07-SEP-2001 (first entry)                                                |
| XX   |                                                                          |
| DE   | Rat glutamate transporter associated protein GRAB4-41 cDNA.              |
| XX   |                                                                          |
| Rat: | glutamate transporter associated protein; GRAB4-41;                      |
| KW   | glutamate transporter; protein; chloride transporter; neurodegeneration; |
| KW   | cytoskeletal stability; nervous system; alcoholism; cholestasis;         |
| KW   | sphingocerebellar ataxia type I; SCAI; GABA metabolism; 85.              |
| OS   | Rattus sp.                                                               |
| XX   |                                                                          |
| XX   |                                                                          |
| XX   | Key Location/Qualifiers                                                  |
| FT   | CDS 1..7167                                                              |
| FT   | /*tag= a                                                                 |
| FT   | /product= "GRAB4-41 protein"                                             |
| FT   | /transl_except= (pos:5142..6150,aa:Val-Ser)                              |
| XX   |                                                                          |
| XX   | W0200130968-f2.                                                          |
| XX   |                                                                          |
| XX   | 03-MAY-2001.                                                             |
| XX   |                                                                          |
| XX   | 23-OCT-2000: 2000MC-US29431.                                             |
| XX   |                                                                          |
| XX   | 23-OCT-1999: 390US-0161007.                                              |
| PR   | 22-MAY-2000: 2000US-0206157.                                             |
| XX   |                                                                          |
| XX   | (UY00 ) UNITI JOHNS HOPKINS SCHOOL MEDICINE.                             |
| XX   |                                                                          |
| XX   | Rothstein JD, Jackson M, Lin G, Law R, Orlov I:                          |
| XX   |                                                                          |
| XX   | WPI: 2001-300498/73.                                                     |
| DR   | P-PSDB: AA001183.                                                        |
| XX   |                                                                          |
| XX   | Novel substantially pure glutamate transporter associated polypeptide    |
| PT   | which modulates intracellular glutamate transport, interacts with        |
| PT   | glutamate transporter protein and has expression pattern in brain        |





30-APR-1997: 9705-0841349.  
(MISHA/) MISHRA L.  
Misha L:  
WPI: 1999-009382/01.  
P-PSDB: AAM81639.  
New isolated early liver development genes - used to develop products for treating, e.g., liver disease, hepatocellular carcinoma, and other pathological disorders, anaemia, ataxia or haemochromatosis  
Claim 2: Fig 2a: 92pp: English.  
This nucleic acid sequence encodes the elf-1 protein (see AAM81639) of the foetal murine liver. The invention provides early coding for developing liver proteins (see AAM81638-42) and embryonic liver cDNA libraries were constructed at days 16.0, 9.11.5, 12.5 and 14.5 post coltus. These were screened with probes derived from growth factors (IGF-I, IGF-II and IGFp-2) and transcriptional activators (C/EBP and LfB1) known to be expressed in the developing liver. Subtractive hybridisation was subsequently performed, and isolated clones were analysed for stage specificity by Southern blotting, sequence, transcript size, abundance and tissue distribution. 3 Elf genes (see AAM81638-42) were identified. The genes were used to screen for the formation of the biliary tree during early liver development. They are used in a claimed method to treat cholestasis, biliary stones, liver obstruction, stricture, primary biliary cirrhosis and primary sclerosing cholangitis. CC Early developing liver proteins and nucleic acids can also be used in the diagnosis of liver diseases and other disorders, including those relating to oncogenesis and tissue repair.  
Sequence 8176 BP: 2349 A; 1850 C; 2147 G; 1830 T; 0 other:  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-750-590-2 x AAM64411 ..  
Align seg 1/1 to: AAM64411 from: 1 to: 8176  
696 LysAsnGlnThrLeuGlnIleLysGluIle 704  
|||||  
4869 AAAAACCCAGCCCTCCACAAAGAAATC 4895  
seq\_name: /SID52/gcgcdata/geneseq/geneseqn/MA1999.DAT:AA13117  
seq\_documentation\_block:  
ID AAX13117 standard: DNA: 31517 BP.  
AC AAX13117:  
XX  
XX 19-MAR-1999 (first entry)  
XX  
XX Enterococcus faecalis genome contig SEQ ID NO:180.  
XX  
XX Enterococcus faecalis: contig: detection: Enterococcal infection:  
XX  
XX vaccine: attenuation: computer readable medium; ds.  
XX  
XX Enterococcus faecalis.  
XX  
XX MO9B50555-A2.  
XX  
XX 12-NOV-1998.  
XX

04-MAY-1998: 98NO-US08985.  
14-NOV-1997: 97US-0066009.  
PR 06-MAY-1997: 97US-0044031.  
PR 16-MAY-1997: 97US-0046555.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Barash SC, Dillon RJ, Kunsch CA:  
XX  
XX WPI: 1999-045171/04.  
XX  
XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
XX  
XX - used to develop products for the detection of Enterococcus and for  
XX  
XX use in vaccines for prevention or attenuation of Enterococcus  
XX  
XX infection.  
XX  
XX Claim 1: Page 986-1002; 208pp: English.  
XX  
XX A computer readable medium has been developed which has recorded on it  
XX  
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
XX  
XX AAX12938 to AAX13919 represent these nucleotide sequences which are  
XX  
XX primary nucleotide sequences, also known as contigs. The computer-based  
XX  
XX system can identify fragments of the Enterococcus faecalis genome in  
XX  
XX commercial importance. The products encoded by these sequences are  
XX  
XX diagnostic for Enterococcus faecalis infection. They can also be used for  
XX  
XX diagnosis of Enterococcal infection in an animal and monitoring  
XX  
XX progression of disease, and for identifying agents which can be used to  
XX  
XX modulate the growth or pathogenicity of Enterococcus faecalis, or  
XX  
XX another related organism, in vivo or in vitro. In particular the  
XX  
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
XX  
XX can be used in vaccines to prevent or attenuate an Enterococcal  
XX  
XX infection.  
XX  
XX Sequence 31517 BP: 10395 A; 5426 C; 6467 G; 9216 T; 13 other:  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-750-590-2 x AAX13117 ..  
Align seg 1/1 to: AAX13117 from: 1 to: 31517  
173 LeuIleAspArgGlyAlaAspIleAsn 181  
|||||  
7492 TTGATTATACGSGGTCCGATATATAT 7518  
seq\_name: /SID52/gcgcdata/geneseq/geneseqn/MA2001.DAT:AA108815  
seq\_documentation\_block:  
ID AA108815 standard: DNA: 228 BP.  
AC AA108815:  
XX  
XX 09-OCT-2001 (first entry)  
XX  
XX  
XX Probe #8805 used to measure gene expression in human breast sample.  
XX  
XX  
XX Probe: human: breast disease; breast cancer: development disorder; ss:  
XX  
XX Inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX  
XX MO200157270-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 29-JAN-2001: 2001WO-US00661.  
XX



```

FT      /tag- C
FT      /label= sticky end
FT      /note= "HindIII"
XX      MO8912678-A.
XX      28-DEC-1989.
XX      16-JUN-1989. 89MO-US02651.
XX      24-JUN-1988. 88US-0210909.
XX      14-DEC-1988. 88US-0284281.
XX      (GETH ) GENENTECH INC.
XX      Liu CC, Miller HI:
XX      WPI: 1990-022538/03.
XX      P-750D; AA003207.
XX      Purified ubiquitin hydrolase and DNA sequences encoding it -
XX      and desired polypeptide prod.
XX      Example 1: Fig 1a: 103bp; English.
XX      This XbaI-HindIII fragment codes for a fusion of human (H2) relaxin
XX      B chain and the amino acid of relaxin B chain. The fusion
XX      protein was expressed by E. coli transformed with the fusion gene.
XX      Inclusion of 35S-labelled Cys in the translation medium resulted in
XX      a radiolabelled fusion protein suitable for use in assays to
XX      investigate cleavage by ubiquitin hydrolase.
XX      See also AA004804 and AA003147.
XX      Sequence 380 BP; 103 A; 83 C; 84 G; 110 T; 0 other:
SQ
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x AA003145/rev ..
Align seg 1/1 to reverse of: AA003145 from: 1 to: 380
693 LeuThrSerLysAsnGlnPheLeu 700
|||||
152 CTTACGACGCAAAATCAACACGCTG 129
seq_name: /SID52/9cgcdata/geneseq/geneseqn/NA1999.DAT.AAV87895
seq_documentation_block:
ID      MAC77203 standard; CDNA. 407 BP.
XX      AAV87895;
XX      12-FEB-1999 (first entry)
XX      EST clone Fg401.
XX      Expressed sequence tag: secreted protein; haematopoietic regulator;
XX      soluble; tumour invasion suppressor; EST; human;
XX      chemokines; chemokinesis; haemostasis; gene therapy; thrombolytics;
XX      receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX      Homo sapiens.
XX      MO9845437-A2.
XX      15-OCT-1998.

```

```

XX      10-APR-1998. 98MO-US06956.
XX      10-APR-1997. 97US-0837312.
XX      (GENY ) GENETICS INSTR INC.
XX      Agostino MJ, Jacobs K, Lysajille ER, McCoy JM, Werberg D;
XX      Racle LA, Spaulding V, Treacy M;
XX      WPI: 1999-070078/06.
XX      New polynucleotides encoding human secreted proteins - derived from
XX      e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX      ovary, pituitary, retina and colon cDNA libraries
XX      Claim 1: Page 215; 641pp; English.
XX      The present sequence represents an expressed sequence tag (EST), and is
XX      a polynucleotide of the invention. The polynucleotides of the invention
XX      are all secreted EST sequences and proteins encoded by a variety of human tissue
XX      sources. The EST sequences and proteins encoded by them are predicted to
XX      have useful biological activities which would make them suitable for
XX      treating, preventing or ameliorating medical conditions and
XX      animals, although not supplying immunological or immune stimulating activities
XX      haematopoietic regulating activity, tissue growth activity,
XX      active/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX      activity, chemokinesis, receptor/ligand activity, anti-inflammatory
XX      activity, cadherin/tumour invasion suppressor activity, tumour inhibition
XX      therapy. The EST sequences are also stated to be useful for gene
XX      therapy.
XX      Sequence 407 BP; 145 A; 88 C; 120 G; 54 T; 0 other:
SQ
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x AAV87895 ..
Align seg 1/1 to: AAV87895 from: 1 to: 407
761 LysLysIleGluGluMetGluLysLeu 768
|||||
48 AAGAACCTTCGAAATGACAAACCTT 71
seq_name: /SID52/9cgcdata/geneseq/geneseqn/NA2000.DAT.MAC77203
seq_documentation_block:
ID      MAC77203 standard; CDNA. 420 BP.
XX      MAC77203;
XX      08-FEB-2001 (first entry)
XX      Human ORFX ORF2758 polynucleotide sequence SEQ ID NO:5515.
XX      Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX      tumour suppressor; antiapoptotic; antiapoptosis; neuroprotective;
XX      anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
XX      antitumour; thrombolytic; coagulant; vasotropic; antidiabetic;
XX      hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX      antineoplastic; antibacterial; antifungal; antipneumatic; antihypoid;
XX      neurodegenerative disorder; cancer; proliferative disorder; hypertension;
XX      cardiovascular disease; diabetes mellitus; hypertension; AIDS;
XX      cholesteryl ester deficiency; malaria; arthropod; infectious; infection;
XX      combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX      allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

```



```

XX AA110455;
DE 12-OCT-2001 (first entry)
XX Probe #388 for gene expression analysis in human cervical cell sample.
XX Probe: human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001: 2001WO-US00670.
XX 04-FEB-2000: 2000US-0180312.
XX 26-MAY-2000: 2000US-0207456.
XX 03-JUN-2000: 2000US-0608408.
XX 03-AUG-2000: 2000US-0632366.
XX 21-SEP-2000: 2000US-0234687.
XX 27-SEP-2000: 2000US-0236359.
XX 04-OCT-2000: 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI: 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX Claim 25: SEQ ID No 388; 487bp; English.
XX PS The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human HeLa cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp://ipo.int/pub/published/Pct_sequences.
XX SO Sequence 471 BP; 133 A; 81 C; 124 G; 133 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x AA110455 ..
Align seg 1/1 to: AA110455 from: 1 to: 471

246 SerAsnysGlyArGclueArp 253
|||||
86 AGCATTAAGCAGNAGACTTGTGG 109

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA131708
seq_documentation_block:
ID AA131708 standard: DNA: 471 BP.
XX AA131708;
XX AC
XX AS
XX OS Homo sapiens.
XX PN 17-OCT-2001 (first entry)

```

```

XX DE Probe #394 used to measure gene expression in human placenta sample.
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001: 2001WO-US00663.
XX 04-FEB-2000: 2000US-0180312.
XX 26-MAY-2000: 2000US-0207456.
XX 03-JUN-2000: 2000US-0608408.
XX 03-AUG-2000: 2000US-0632366.
XX 21-SEP-2000: 2000US-0234687.
XX 27-SEP-2000: 2000US-0236359.
XX 04-OCT-2000: 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI: 2001-488997/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX Claim 25: SEQ ID No 394; 654bp; English.
XX PS The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SO Sequence 471 BP; 133 A; 81 C; 124 G; 133 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x AA131708 ..
Align seg 1/1 to: AA131708 from: 1 to: 471

246 SerAsnysGlyArGclueArp 253
|||||
86 AGCATTAAGCAGNAGACTTGTGG 109

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA100391
seq_documentation_block:
ID AA100391 standard: DNA: 471 BP.
XX AA100391;
XX AC
XX AS
XX OS Homo sapiens.
XX PN 09-OCT-2001 (first entry)
XX DE Probe #382 used to measure gene expression in human breast sample.
XX Probe: human; breast disease; breast cancer; development disorder; ss.
XX KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.

```

XX 09-AUG-2001.  
 XX 29-JAN-2001: 2001MO-US00661.  
 XX 04-FEB-2000: 2000OUS-0180312.  
 XX 26-MAY-2000: 2000OUS-0207456.  
 XX 30-JUN-2000: 2000OUS-0608408.  
 XX 03-AUG-2000: 2000OUS-0632366.  
 XX 21-SEP-2000: 2000OUS-0234687.  
 XX 27-SEP-2000: 2000OUS-0236359.  
 XX 04-OCT-2000: 2000OEN-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR:  
 PI WPI: 2001-476286/51.  
 DR WPI: 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 XX in a human breast.  
 XX Claim 25: SEQ ID NO 382: 322pp: English.  
 XX The present invention relates to novel single exon nucleic acid probes.  
 XX The present sequence is one such probe. The probes are useful for  
 XX measuring human gene expression in a human breast sample, where the probe  
 XX hybridizes at high stringency to a nucleic acid expressed in the human  
 XX breast. The probes are useful for predicting, diagnosing, grading,  
 XX particularly the presence of the human breast,  
 XX including the presence of the human breast, inflammatory diseases  
 XX of the breast, fibrocystic changes, proliferative breast disease and  
 XX non-carcinoma tumors.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX Sequence 471 BP: 133 A: 81 C: 124 G: 133 T: 0 other:  
 SO

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x MA100391 ..  
 Align seg 1/1 to: MA100391 from: 1 to: 471

246 SerAnlysglyAAGtGtLeuTRP 253  
 86 AGCAATTAACGACAACTCTCTCG 109  
 seq\_name: /SID52/gcgdata/geneseq/geneseq/MA2000.DMT:MAC94987  
 seq\_documentation\_block:  
 ID AAC94987 standard: cDNA: 540 BP.  
 AC AAC94987:  
 XX 19-FEB-2001 (first entry)  
 XX Cat flea hindgut and Malpighian tubule (HMT) cDNA, S50 ID NO:1482.  
 XX Cat flea hindgut and Malpighian tubule nucleic acid: HMT:  
 XX Flea infestation: vaccine: antiparasitic: therapeutic target:  
 XX diagnosis: detection: ss.  
 XX Ctenocephalides felis.  
 XX MO200061621-A2.

XX 19-OCT-2000.  
 XX 07-APR-2000: 2000MO-US09437.  
 XX 09-APR-1999: 99US-0128704.  
 XX (HESK-) HESKA CORP.  
 XX Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N:  
 XX WPI: 2000-656323/63.  
 DR WPI: 2000-656323/63.  
 XX Plea Malpighian tubule and head and nerve cord tissue derived nucleic  
 XX acid useful for the prevention, diagnosis and treatment of flea  
 XX infestations.  
 XX Claim 26: Page 730-731: 964pp: English.  
 XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic  
 XX acids which are expressed in hindgut (HMT) and Malpighian tubule (MT) tissue  
 XX or head and nerve cord (HNC) tissue. The invention additionally encompasses expression  
 XX encoded proteins. The invention additionally encompasses expression the  
 XX constructs, recombinant viruses and recombinant cells comprising the  
 XX nucleic acids of the invention, recombinant production of the proteins,  
 XX antibodies against the proteins, a method of identifying inhibitors of  
 XX the proteins, and compositions comprising the inhibitors for  
 XX administering to a host animal.  
 XX The invention also includes the nucleic acids, and the proteins they  
 XX encode may be used in the diagnosis and treatment of diseases  
 XX associated with flea infestations. For example, the  
 XX used to produce an HMT or HNC protein according to standard recombinant  
 XX DNA methodology by inserting the nucleic acids into a host cell and  
 XX culturing the cell to express the protein. The HMT and HNC nucleic acids  
 XX may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect  
 XX sequences in the presence of cat flea or other homologous nucleic acid  
 XX sequences. In addition, the nucleic acids may be used to study the expression and  
 XX function of the proteins and help to study the production of specific  
 XX proteins may be used as antigens in the production of specific  
 XX antibodies, and in assays to identify modulators (agonists and  
 XX antagonists) of HMT and/or HNC protein expression and activity. The  
 XX anti-HMT/HNC protein antibodies and antagonists may also be used to  
 XX downregulate protein expression and activity. The antibodies may also be  
 XX used as diagnostic agents for detecting the presence of flea polypeptides  
 XX (e.g., of enzymes linked immunosorbent assay (ELISA)). The  
 XX present sequence represents a cat flea HMT cDNA of the invention.  
 XX Sequence 540 BP: 198 A: 85 C: 101 G: 155 T: 1 other:  
 SO

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x AAC94987/rev ..  
 Align seg 1/1 to reverse of: AAC94987 from: 1 to: 540

693 LeuThSerLysAngGlnThrLeu 700  
 288 CTTCACGCAAAACAAACAGCTTG 265  
 seq\_name: /SID52/gcgdata/geneseq/geneseq/MA2000.DMT:MAC35479  
 seq\_documentation\_block:  
 ID AAC35479 standard: DNA: 551 BP.  
 AC AAC35479:  
 XX 17-OCT-2000 (first entry)  
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 10346.

XX Hybridisation assay: genetic mapping; gene expression control;  
XX Protein identification: signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 02-APR-1999; 99US-0128232.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134221.  
XX 18-MAY-1999; 99US-0134370.  
XX 19-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 27-MAY-1999; 99US-0136022.  
XX 28-MAY-1999; 99US-0136722.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138610.  
XX 16-JUN-1999; 99US-0138452.  
XX 16-JUN-1999; 99US-0138453.  
XX 17-JUN-1999; 99US-0138454.  
XX 18-JUN-1999; 99US-0138455.  
XX 18-JUN-1999; 99US-0138456.  
XX 18-JUN-1999; 99US-0138457.  
XX 18-JUN-1999; 99US-0138458.  
XX 18-JUN-1999; 99US-0138459.  
XX 18-JUN-1999; 99US-0138460.  
XX 18-JUN-1999; 99US-0138461.  
XX 18-JUN-1999; 99US-0138462.  
XX 18-JUN-1999; 99US-0138463.  
XX 18-JUN-1999; 99US-0138464.  
XX 18-JUN-1999; 99US-0138465.  
XX 18-JUN-1999; 99US-0138466.  
XX 18-JUN-1999; 99US-0138467.  
XX 18-JUN-1999; 99US-0138468.  
XX 18-JUN-1999; 99US-0138469.  
XX 21-JUN-1999; 99US-0138909.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0141843.  
PR 02-JUL-1999; 99US-0141844.  
PR 02-JUL-1999; 99US-0141845.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144096.  
PR 18-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145143.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0145388.  
PR 02-AUG-1999; 99US-0145388.  
PR 02-AUG-1999; 99US-0145388.  
PR 02-AUG-1999; 99US-0145388.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148310.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 16-AUG-1999; 99US-0149175.  
PR 16-AUG-1999; 99US-0149175.  
PR 16-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149929.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151503.

```

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157652.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159320.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161320.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores:
  Quality: 8.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Percent Identity: 100.000

```

```

align seg 1/1 to reverse of: AAC35479 from: 1 to: 551
US-09-750-590-2 x AAC35479/rev ..

```

```

1146 SeqVblProblemaIsclmLstLen 1153
|||||
229 TCAGTACCACTGGCGACACCTTA 206

```

```

seq_name: //SID52/gcgcdata/geneseq/geneseq/NA2001.DAT:AA116572

```

```

seq_documentation_block:

```

```

ID AA116572 standard: DNA; 570 BP.

```

```

AC AA116572.

```

```

DT 12-OCT-2001 (first entry)

```

```

DE Probe #505 for gene expression analysis in human cervical cell sample.
XX
XX

```

```

KY Probe: human: microarray: gene expression: cervical epithelial cell;
KW Cervical cancer; 35.
XX
XX Homo sapiens.
PN WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632666.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000US-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn Sq, Hanzel DK, Chan W, Rank DR;
PI WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX Claim 25; SEQ ID No 6505; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX CC. Note: The sequence data for this probe is derived from a part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 570 BP; 137 A; 150 C; 138 G; 145 T; 0 other;
S0

```

```

alignment_scores:
  Quality: 8.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Percent Identity: 100.000

```

```

align seg 1/1 to reverse of: AA116572 from: 1 to: 570
US-09-750-590-2 x AA116572/rev ..

```

```

1334 ArgGincInclmLstLen 1341
|||||
403 AGCGAACCACGCTCATGATCT 380

```

```

seq_name: //SID52/gcgcdata/geneseq/geneseq/NA2001.DAT:AA139649

```

```

seq_documentation_block:

```

```

ID AA139649 standard: DNA; 570 BP.

```

```

AC AA139649.

```

```

DT 17-OCT-2001 (first entry)

```

```

DE Probe #8335 used to measure gene expression in human placenta sample.
XX
XX

```

```

KW Probe: microarray: human: placenta: antenatal diagnosis;
XX
XX genetic disorder; ss.
XX
XX Homo sapiens.
OS

```



```

XX  MO200157272-A2.
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001: 2001MO-US00663.
XX
XX  04-FEB-2000: 2000US-0180312.
XX  26-MAY-2000: 2000US-0207456.
XX  30-JUN-2000: 2000US-0608408.
XX  03-AUG-2000: 2000US-0632485.
XX  23-SEP-2000: 2000US-0234285.
XX  27-SEP-2000: 2000US-0236359.
XX  04-OCT-2000: 2000CB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR:
XX  WPI: 2001-488897/53.
XX  Human genome-derived single exon nucleic acid probes useful for
XX  analyzing gene expression in human placenta -
XX
XX  Claim 25: SEQ ID No 8335; 654pp; English.
XX
XX  The present invention relates to single exon nucleic acid probes (SENP).
XX  The present sequence is one specific probe, measuring and displaying gene
XX  expression in samples derived from human placenta. The probes are useful
XX  for antenatal diagnosis of human genetic disorders.
XX
XX  Sequence 570 BP: 137 A; 150 C; 138 G; 145 T; 0 other;
XX
XX  alignment_scores:
XX      Quality: 8.00      Length: 8
XX      Ratio: 1.000      Gaps: 0
XX  Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX  alignment_block:
XX  US-09-750-590-2 x AA139649/rev ..
XX
XX  Align seg 1/1 to reverse of: AA139649 from: 1 to: 570
XX
XX  1334 ATGCTGTCGCTCCGCGAGGACGAGC 1341
XX  403 AGCCAAAGCCAGCATTCATCTCT 380
XX
XX  seq_name: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:AAF22599
XX
XX  seq_documentation_block:
XX  ID AAF22599 standard: cDNA: 657 BP.
XX
XX  AAF22599:
XX
XX  DT 26-MAR-2001 (first entry)
XX
XX  DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:178.
XX
XX  KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX  cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
XX  OS Homo sapiens.
XX
XX  PN MO200073801-A2.
XX
XX  PD 07-BEC-2000.
XX
XX  PF 26-MAY-2000: 2000MO-US14749.
XX  28-MAY-1999: 99US-0136526.
XX  10-SEP-1999: 99US-0153454.
XX

```

```

XX  (LUDM-) LUDWIG INST CANCER RES.
XX
XX  Obata Y:
XX
XX  WPI: 2001-025274/03.
XX
XX  Nucleic acids encoding breast, gastric and prostate cancer associated
XX  antigen precursors, useful for diagnosing and treating cancer
XX  characterized by expression of an abnormal amount of a protein, e.g.
XX  cancer.
XX
XX  Claim 50: Page 332; 799pp; English.
XX
XX  AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX  represent nucleotide sequences encoding human breast, gastric and
XX  prostate cancer associated antigen precursors. (AA631722 to AA631970
XX  AAF23233 to AAF23444) encode human breast, gastric and prostate CAP protein
XX  respectively. CAPs have cytostatic activity and can be used in the
XX  production of cancer vaccines. The human CAP proteins, peptides, nucleic
XX  acids or anti-CAP antibodies are useful for diagnosing and treating a
XX  condition characterized by expression of an abnormal amount of a protein,
XX  e.g. cancer.
XX
XX  Sequence 657 BP: 123 A; 184 C; 261 G; 87 T; 2 other;
XX
XX  alignment_scores:
XX      Quality: 8.00      Length: 8
XX      Ratio: 1.000      Gaps: 0
XX  Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX  alignment_block:
XX  US-09-750-590-2 x AAF22599 ..
XX
XX  Align seg 1/1 to: AAF22599 from: 1 to: 657
XX
XX  1168 AATGATGATGCTGCTCCGCGAGGACGAGC 224
XX  201 GCTTCTGCTCCGCGAGGACGAGC 224
XX
XX  seq_name: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:AAH06793
XX
XX  seq_documentation_block:
XX  ID AAH06793 standard: cDNA: 720 BP.
XX
XX  AAH06793:
XX
XX  DT 26-JUN-2001 (first entry)
XX
XX  DE Human cDNA clone (5'-primer) SEQ ID NO:3628.
XX
XX  KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX  OS Homo sapiens.
XX
XX  PN EP1074617-A2.
XX
XX  PD 07-FEB-2001.
XX
XX  DE 28-JUL-2000: 2000EP-0116126.
XX
XX  29-JUL-1999: 99UP-0248036.
XX  27-AUG-1999: 99UP-0300253.
XX  11-JAN-2000: 2000JP-0118776.
XX  02-MAY-2000: 2000JP-0183767.
XX  09-JUN-2000: 2000JP-0241899.
XX
XX  (HELI-) HELIX RES INST.
XX
XX  Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J:
XX  Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T:
XX

```

```

XX      WPI: 2001-318749/34.
XX
XX      Primer sets for synthesizing polynucleotides, particularly the 5602
XX      full-length cDNAs defined in the specification, and for the detection
XX      of the abnormality of the proteins encoded by the
XX      full-length cDNAs -
XX
XX      Claim 1: SEQ ID 3628; 2537bp + CD ROM; English.
XX
XX      The present invention describes primer sets for synthesizing 5602
XX      full-length cDNAs defined in the specification. Where a primer set
XX      comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX      to the 5602 complementary strand of a polynucleotide which comprises one of
XX      the 5602 nucleotide sequences defined in the specification, where the
XX      oligonucleotide comprises a sequence complementary to a 5'-end
XX      of an oligonucleotide comprising a sequence complementary to a 5'-end
XX      complementary strand of a polynucleotide which comprises a 5'-end
XX      sequence and an oligonucleotide comprising a sequence complementary to a
XX      polynucleotide which comprises a 3'-end sequence, where the
XX      oligonucleotide comprises at least 15 nucleotides and the combination of
XX      the 5'-end sequence/3'-end sequence is selected from those defined in
XX      the specification. The primer sets can be used in antisense therapy and
XX      particularly full-length cDNAs. The primers are also useful for the
XX      detection and/or diagnosis of the abnormality of the proteins encoded by
XX      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX      AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX      AAH95893 represent human amino acid sequences; and AAH15629 to AAH13632
XX      represent oligonucleotides, all of which are used in the exemplification
XX      of the present invention.
XX
XX      Sequence 720 BP; 203 A; 119 C; 175 G; 215 T; 8 other;
XX
XX      alignment_scores:
XX          Quality:      8.00      Length:      8
XX          Ratio:        1.000      Gaps:        0
XX      Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX      alignment_block:
XX      US-09-750-590-2 x AAH06793/rev
XX
XX      Align seg 1/1 to reverse of: AAH06793 from: 1 to: 720
XX
XX      799 LeuylSerAnilletHrgIuLeu 806
XX      |||||||
XX      72 CTGAAGTCGACACATCAGCAACTG 49
XX
XX      seq_name: /SIDS2/gcdata/geneseq/geneseq/NA2001.DAT:AAH06245
XX
XX      seq_documentation_block:
XX      ID: AAH06245 standard: cDNA: 747 BP.
XX
XX      AAH06245:
XX      |||||||
XX      26-JUN-2001 (first entry)
XX
XX      Human cDNA clone (5'-primer) SEQ ID NO:3080.
XX
XX      Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX      Homo sapiens.
XX
XX      BP1074617-A2.
XX
XX      PD 07-FEB-2001.
XX
XX      PF 28-JUL-2000; 2000EP-0116126.
XX
XX      PR 29-JUL-1999; 99JP-0248036.
XX
XX      PR 27-AUG-1999; 99JP-0300253.

```

```

PR      11-JAN-2000; 2000JP-0118776.
PR      02-JAN-2000; 2000JP-0103367.
PR      09-JUN-2000; 2000JP-0241859.
XX
XX      (HELI-) HELIX RES INST.
XX
XX      P1 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX      P1 Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX      WPI: 2001-318749/34.
XX
XX      Primer sets for synthesizing polynucleotides, particularly the 5602
XX      full-length cDNAs defined in the specification, and for the detection
XX      of the abnormality of the proteins encoded by the
XX      full-length cDNAs -
XX
XX      Claim 1: SEQ ID 3080; 2537bp + CD ROM; English.
XX
XX      The present invention describes primer sets for synthesizing 5602
XX      full-length cDNAs defined in the specification. Where a primer set
XX      comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX      to the 5602 complementary strand of a polynucleotide which comprises one of
XX      the 5602 nucleotide sequences defined in the specification, where the
XX      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX      of an oligonucleotide comprising a sequence complementary to the
XX      complementary strand of a polynucleotide which comprises a 5'-end
XX      sequence and an oligonucleotide comprising a sequence complementary to a
XX      polynucleotide which comprises a 3'-end sequence, where the
XX      oligonucleotide comprises at least 15 nucleotides and the combination of
XX      the 5'-end sequence/3'-end sequence is selected from those defined in
XX      the specification. The primer sets can be used in antisense therapy and
XX      particularly full-length cDNAs. The primers are also useful for the
XX      detection and/or diagnosis of the abnormality of the proteins encoded by
XX      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX      AAH95893 represent human cDNA sequences; AAH92446 to
XX      AAH95893 represent human amino acid sequences; and AAH15629 to AAH13632
XX      represent oligonucleotides, all of which are used in the exemplification
XX      of the present invention.
XX
XX      Sequence 747 BP; 221 A; 129 C; 148 G; 240 T; 9 other;
XX
XX      alignment_scores:
XX          Quality:      8.00      Length:      8
XX          Ratio:        1.000      Gaps:        0
XX      Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX      alignment_block:
XX      US-09-750-590-2 x AAH06245/rev
XX
XX      Align seg 1/1 to reverse of: AAH06245 from: 1 to: 747
XX
XX      799 LeuylSerAnilletHrgIuLeu 806
XX      |||||||
XX      343 CTGAAGTCGACACATCAGCAACTG 320

```



**THIS PAGE BLANK (USPTO)**



```

seq_name: /cgn2_6/prodata/2/lna/5b_COMB.seq:US-08-841-349-3
seq_documentation_block:
: Sequence 3, Application: US/08841349B
: Patent No. 5953594
: GENERAL INFORMATION:
: APPLICANT: MEZES, PETER S
: APPLICANT: KAPLAN, DONALD A
: TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
: FILE REFERENCE: XX/PO4470USO
: CURRENT APPLICATION NUMBER: US/08/841,349B
: CURRENT FILING DATE: 1997-04-30
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 3
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (333)..(6794)
US-08-841-349-3

```

```

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

align seg 1/1 to: US-08-841-349-3 ..
US-09-750-590-2 x US-08-841-349-3

```

```

596 LysAaenclThrlaenclnlysgluile 704
|||||
4869 AAAACCAAGACCCCTCCAGAAAGAAATC 4895

```

```

seq_name: /cgn2_6/prodata/2/lna/5b_COMB.seq:US-08-841-349-5

```

```

seq_documentation_block:
: Sequence 5, Application: US/08841349B
: Patent No. 5953594
: GENERAL INFORMATION:
: APPLICANT: MEZES, PETER S
: APPLICANT: KAPLAN, DONALD A
: TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
: FILE REFERENCE: XX/PO4470USO
: CURRENT APPLICATION NUMBER: US/08/841,349B
: CURRENT FILING DATE: 1997-04-30
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 5
: TYPE: DNA
: ORGANISM: Mus musculus
US-08-841-349-5

```

```

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

align seg 1/1 to: US-08-841-349-5 ..
US-09-750-590-2 x US-08-841-349-5

```

```

696 LysAaenclThrlaenclnlysgluile 704
|||||
4869 AAAACCAAGACCCCTCCAGAAAGAAATC 4895

```

```

seq_name: /cgn2_6/prodata/2/lna/5b_COMB.seq:US-08-822-028-26

```

```

seq_documentation_block:
: Sequence 26, Application: US/08822028
: Patent No. 5993813
: GENERAL INFORMATION:
: APPLICANT: MEZES, PETER S
: APPLICANT: GOURLE, BRIAN B
: APPLICANT: ANDERSON, MARK W
: APPLICANT: KAPLAN, DONALD A
: APPLICANT: RIXON, BRIAN W
: TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,
: MODIFIED ANTIBODIES FOR CANCER TREATMENT
: NUMBER OF SEQUENCES: 74
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DUANE C. ULMER
: STREET: MIDLAND
: STATE: MICHIGAN
: COUNTRY: USA
: ZIP: 48641-1967
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION NUMBER: US/08/822,028
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA: US 08/040,687
: APPLICATION NUMBER: US 08/040,687
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: ULMER, DUANE C
: REFERENCE NUMBER: 34,941
: REFERENCE/DOCKET NUMBER: C-37,075C
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (517) 636-8104
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 816 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: ORIENTATION: 5' to 3'
: MOLECULE TYPE: DNA (genomic)
US-08-822-028-26

```

```

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

align seg 1/1 to reverse of: US-08-822-028-26/rev ..
US-09-750-590-2 x US-08-822-028-26/rev

```

```

1056 LysAaenclnlysgluile 1063
|||||
772 MGGATCTCTCAAGAAATCACT 749

```

```

seq_name: /cgn2_6/prodata/2/lna/5b_COMB.seq:US-08-479-285-26

```

```

seq_documentation_block:
: Sequence 26, Application: US/08479285
: Patent No. 6207815
: GENERAL INFORMATION:
: APPLICANT: MEZES, PETER S
: APPLICANT: GOURLE, BRIAN B
: APPLICANT: RIXON, MARK W
: APPLICANT: ANDERSON, MARK W
: APPLICANT: KAPLAN, DONALD A

```

```

1  APPLICANT: SCHOLOM, JEFFREY
2  TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,
3  TITLE OF INVENTION: MODIFIED ANTIBODIES FOR CANCER TREATMENT
4  NUMBER OF SEQUENCES: 74
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: DUANE C. ULMER
7  STREET: P. O. BOX 1967
8  CITY: MIDLAND
9  STATE: MICHIGAN
10 COUNTRY: USA
11 ZIP: 48641-1967
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE RELEASE: PC-DOS/MS-DOS
16 CURRENT APPLICATION DATA:
17 FILING DATE: 07-JUN-1995
18 CLASSIFICATION: 536
19 APPLICATION NUMBER: US/08/479,265
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/040687
22 FILING DATE: 31-MAR-1993
23 ATTORNEY/AGENT INFORMATION:
24 NAME: ULMER, DUANE C.
25 REGISTRATION NUMBER: C-34,941
26 REFERENCE/DOCKET NUMBER: C-37,075C
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (517) 636-8104
29 INFORMATION FOR SEQ ID NO: 26:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 816 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 US-08-479-285-26

```

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment block:

US-09-750-590-2 x US-08-479-285-26/rev ..

Align seg 1/1 to reverse of: US-08-479-285-26 from: 1 to: 816

1056 LysAspIeuLengInIyTyThr 1063

|||||

772 AAGGATCTTACGAGGAAATCACT 749

seq\_name: /cgn2\_6/ptcdat2/1na/5b.CONB.seq:US-08-822-028-5

seq\_documentation\_block:

```

1  Sequence 5, Application US/08822028
2  Patent No. 593813
3  GENERAL INFORMATION:
4  APPLICANT: MEZES, PETER S
5  APPLICANT: SCHOLOM, JEFFREY
6  APPLICANT: RIXON, MARK B
7  APPLICANT: ANDERSON, WM KERR
8  APPLICANT: KAPLAN, DONALD A
9  APPLICANT: SCHOLOM, JEFFREY
10 TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,
11 TITLE OF INVENTION: MODIFIED ANTIBODIES FOR CANCER TREATMENT
12 NUMBER OF SEQUENCES: 74
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: DUANE C. ULMER
15 STREET: P. O. BOX 1967
16 CITY: MIDLAND
17 STATE: MICHIGAN

```

```

1  COUNTRY: USA
2  ZIP: 48641-1967
3  COMPUTER READABLE FORM:
4  MEDIUM TYPE: Floppy disk
5  OPERATING SYSTEM: PC-DOS/MS-DOS
6  SOFTWARE: Patent Release #1.0, Version #1.25
7  CURRENT APPLICATION DATA:
8  FILING DATE:
9  CLASSIFICATION:
10 APPLICATION NUMBER: US/08/822,028
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER: US 08/040,687
13 FILING DATE:
14 ATTORNEY/AGENT INFORMATION:
15 NAME: ULMER, DUANE C.
16 REGISTRATION NUMBER: C-34,941
17 REFERENCE/DOCKET NUMBER: C-37,075C
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (517) 636-8104
20 INFORMATION FOR SEQ ID NO: 5:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 862 base pairs
23 TYPE: nucleic acid
24 STRANDEDNESS: double
25 TOPOLOGY: linear
26 MOLECULE TYPE: DNA (genomic)
27 FEATURE: CDS
28 NAME/KEY: j01n(384..429, 509..862)
29 LOCATION: j01n(384..429, 509..519)
30 NAME/KEY: mat-peptide
31 LOCATION: 320..862
32 US-08-822-028-5

```

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment block:

US-09-750-590-2 x US-08-822-028-5/rev ..

Align seg 1/1 to reverse of: US-08-822-028-5 from: 1 to: 862

1056 LysAspIeuLengInIyTyThr 1063

|||||

818 AAGGATCTTACGAGGAAATCACT 795

seq\_name: /cgn2\_6/ptcdat2/1na/5b.CONB.seq:US-08-479-285-5

seq\_documentation\_block:

```

1  Sequence 5, Application US/08479285
2  Patent No. 6207815
3  GENERAL INFORMATION:
4  APPLICANT: MEZES, PETER S
5  APPLICANT: SCHOLOM, JEFFREY
6  APPLICANT: RIXON, MARK B
7  APPLICANT: ANDERSON, WM KERR
8  APPLICANT: KAPLAN, DONALD A
9  APPLICANT: SCHOLOM, JEFFREY
10 TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,
11 TITLE OF INVENTION: MODIFIED ANTIBODIES FOR CANCER TREATMENT
12 NUMBER OF SEQUENCES: 74
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: DUANE C. ULMER
15 STREET: P. O. BOX 1967
16 CITY: MIDLAND
17 STATE: MICHIGAN

```

```

? COUNTRY: USA
? ZIP: 48641-1667
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 536
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/419,285
? FILING DATE: 31-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: LEBEN, DANIE C. 34,941
? REGISTRATION NUMBER: 37,459
? TELEPHONE: (517) 636-8104
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 862 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: join(384..429, 509..862)
? NAME/KEY: sig_peptide
? LOCATION: join(384..429, 509..519)
? NAME/KEY: mat_peptide
? LOCATION: 520..862
US-08-479-285-5

```

```

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x US-08-479-285-5/rev ..
Align seg 1/1 to: reverse of: US-08-479-285-5 from: 1 to: 862

```

```

1056 lysaapleuileuglnlystythr 1063
|||||
818 AACGACTCTGACAAATACGACT 795
seq_name: /cgn2_6/prodata/2/lna/5A_CONB.seq:US-09-031-485-40
seq_documentation_block:
? Sequence 40, Application US/09031485
? Patent No. 5824306
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? APPLICANT: Blehm, E. Scott
? TITLE OF INVENTION: DIRECTIONAL AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Versar, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525

```

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? FILING DATE: US/09/031,485
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/847,429
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Versar, Carol Talkington
? REGISTRATION NUMBER: 37,459
? TELEPHONE: 970/483-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 40:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 906 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? MOLECULE TYPE: linear
US-09-031-485-40

```

```

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x US-09-031-485-40 ..
Align seg 1/1 to: US-09-031-485-40 from: 1 to: 906

```

```

92 leuailsluailaialystrygly 99
|||||
625 TTGCATTGACTGCTAAATACGGCC 648
seq_name: /cgn2_6/prodata/2/lna/5A_CONB.seq:US-09-031-485-41
seq_documentation_block:
? Sequence 41, Application US/09031485
? Patent No. 5824306
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? APPLICANT: Blehm, E. Scott
? TITLE OF INVENTION: DIRECTIONAL AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Versar, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? FILING DATE: US/09/031,485
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:

```



APPLICATION NUMBER: US/08/847,429  
 FILING DATE: 24-APR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verser, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 REFERENCE/DOCKET NUMBER: HW-5  
 TELEPHONE: 970/484-9505  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 41:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 906 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-09-031-485-41

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x US-09-031-485-41/rev ..

Align seg 1/1 to reverse of: US-09-031-485-41 from: 1 to: 906

92 LeuHLSLenaLaLaLaTyTYcLY 99  
 |||||||||||||||||||||||||  
 282 TTCGATTAGCTGCTAATACGCG 259

seq\_name: /cgn2\_6/prodata/2/lna/5a\_LCOMB.seq:US-08-847-429A-40

seq\_documentation\_block:

Sequence 40, Application US/08847429A

Patent No. 5827692

GENERAL INFORMATION:

APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scot

TITLE OF INVENTION: DISFILARIA AND BRUGIA ANKYRIN

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/847,429A

FILING DATE: 24-APR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: HW-5

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 906 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-847-429A-40

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x US-08-847-429A-40 ..

Align seg 1/1 to: US-08-847-429A-40 from: 1 to: 906

92 LeuHLSLenaLaLaLaTyTYcLY 99  
 |||||||||||||||||||||||||  
 625 TTCGATTAGCTGCTAATACGCG 648

seq\_name: /cgn2\_6/prodata/2/lna/5a\_LCOMB.seq:US-08-847-429A-41

seq\_documentation\_block:

Sequence 41, Application US/08847429A

Patent No. 5827692

GENERAL INFORMATION:

APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scot

TITLE OF INVENTION: DISFILARIA AND BRUGIA ANKYRIN

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/847,429A

FILING DATE: 24-APR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: HW-5

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 41:

SEQUENCE CHARACTERISTICS:

LENGTH: 906 nucleotides

STRANDEDNESS: single

TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-847-429A-41

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x US-08-847-429A-41/rev ..



```

? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/031-485
? FILING DATE:
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/847-429
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37/459
? REFERENCE/DOCKET NUMBER: HW-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 37:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 908 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..906
? US-09-031-485-37

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
? US-09-750-590-2 x US-09-031-485-37 ..
?
? Align seq 1/1 to: US-09-031-485-37 from: 1 to: 908
?
? 92 LeuHISLeuAlaAlaValSTyrcly 99
? |||||
? 625 TTGCATTAGCTCTTAATACGCG 648
?
seq_name: /cgn2.6/prodata2/1na/5A_COMB.seq:US-09-031-485-39
seq_documentation_block:
? Sequence 39, Application US/09031485
? Patent No. 5843062
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? ADDRESS: Blehm, E. Scott
? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKRYIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/847-429
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37/459
? REFERENCE/DOCKET NUMBER: HW-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? CURRENT APPLICATION DATA:

```

```

? APPLICATION NUMBER: US/09/031-485
? FILING DATE:
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/847-429
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37/459
? REFERENCE/DOCKET NUMBER: HW-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 39:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 908 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-031-485-39

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
? US-09-750-590-2 x US-09-031-485-39/rev ..
?
? Align seq 1/1 to reverse of: US-09-031-485-39 from: 1 to: 908
?
? 92 LeuHISLeuAlaAlaValSTyrcly 99
? |||||
? 284 TTGCATTAGCTCTTAATACGCG 261
?
seq_name: /cgn2.6/prodata2/1na/5A_COMB.seq:US-08-847-429A-37
seq_documentation_block:
? Sequence 37, Application US/08847429A
? Patent No. 5827692
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? ADDRESS: Blehm, E. Scott
? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKRYIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/847-429A
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37/459
? REFERENCE/DOCKET NUMBER: HW-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505

```

```

? INFORMATION FOR SEQ ID NO: 37:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 908 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..906
? US-08-847-429A-37

```

```

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-750-590-2 x US-08-847-429A-37
Align seg 1/1 to: US-08-847-429A-37 from: 1 to: 908

```

```

92 LeuHAsleuAlaAlaLysTyrGly 99
|||||
625 TTGCATTAGCTGCTAAATACGCG 648

```

```

seq_name: /cgn2_6/plodact/2/lna/5A_COMB.seq:US-08-847-429A-39
seq_documentation_block:
? Sequence 39, Application US/08847429A
? Patent No. 5827692
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? INVENTOR: Tang, Liang; Blum, E. Scott
? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/847,429A
? FILING DATE: 24-APR-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37,459
? REFERENCE/DOCKET NUMBER: HW-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/493-5071
? INFORMATION FOR SEQ ID NO: 39:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 908 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-08-847-429A-39

```

```

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-750-590-2 x US-08-847-429A-39/rev
Align seg 1/1 to reverse of: US-08-847-429A-39 from: 1 to: 908

```

```

92 LeuHAsleuAlaAlaLysTyrGly 99
|||||
284 TTGCATTAGCTGCTAAATACGCG 261

```

```

seq_name: /cgn2_6/plodact/2/lna/6A_COMB.seq:US-09-065-474-37
seq_documentation_block:
? Sequence 37, Application US/09065474
? Patent No. 5827692
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? INVENTOR: Tang, Liang; Blum, E. Scott
? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 171
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/065,474
? FILING DATE: 24-APR-1998
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 37,459
? REFERENCE/DOCKET NUMBER: HW-5-C1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/494-9505
? INFORMATION FOR SEQ ID NO: 37:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 908 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..906
? US-09-065-474-37

```

```

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-750-590-2 x US-09-065-474-37
Align seg 1/1 to: US-09-065-474-37 from: 1 to: 908

```

92 LeuHISLeuAlaIaLysTyrGly 99  
 |||||||  
 625 TTGCATTAGCTCTCAATAGTACGCC 648

seq\_name: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:US-09-065-474-39

seq\_documentation\_block:

Sequence 39, Application US/09065474  
 Patent No. 6063599  
 GENERAL INFORMATION:  
 APPLICANT: Tanq, Liang  
 APPLICANT: Blehm, E. Scot  
 TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 TITLE OF INVENTION: US\$ THEREOF  
 NUMBER OF SEQUENCES: 111  
 CORRESPONDENCE ADDRESS: 111  
 ADDRESSER: Carol Talkington Verser, Ph.D.  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80525  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: WordPerfect for Windows, Version 7.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/065,474  
 FILING DATE: 24-APR-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Versey, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 REFERENCE/DOCKET NUMBER: HM-5-C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/493-7272  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 908 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-09-065-474-39

alignment\_scores:

Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x US-09-065-474-39/rev ..

Align seg 1/1 to reverse of: US-09-065-474-39 from: 1 to: 908

92 LeuHISLeuAlaIaLysTyrGly 99  
 |||||||

284 TTGCATTAGCTCTCAATAGTACGCC 261

seq\_documentation\_block:

Sequence 25, Application US/09031485  
 Patent No. 5824306  
 GENERAL INFORMATION:  
 APPLICANT: Tanq, Liang  
 APPLICANT: Blehm, E. Scot  
 TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 TITLE OF INVENTION: US\$ THEREOF  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS: 85  
 ADDRESSER: Carol Talkington Verser, Ph.D.  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 TITLE OF INVENTION: US\$ THEREOF  
 NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS: 85

ADDRESSER: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/031,485

FILING DATE: 24-APR-1997

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/947,429

FILING DATE: 24-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Versey, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: HM-5

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 909 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-09-031-485-25

alignment\_scores:

Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x US-09-031-485-25 ..

Align seg 1/1 to: US-09-031-485-25 from: 1 to: 909

92 LeuHISLeuAlaIaLysTyrGly 99  
 |||||||

628 TTGCATTAGCTCTCAATAGTACGCC 651

seq\_documentation\_block:

Sequence 26, Application US/09031485  
 Patent No. 5824306  
 GENERAL INFORMATION:  
 APPLICANT: Tanq, Liang  
 APPLICANT: Blehm, E. Scot  
 TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 TITLE OF INVENTION: US\$ THEREOF  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS: 85  
 ADDRESSER: Carol Talkington Verser, Ph.D.  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA

ZIP: 80525  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: Wordperfect for Windows, Version 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 05/09/031.485  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/847.429  
FILING DATE: 24-Apr-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: HW-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 970/484-9505  
TELEFAX: 970/484-9505  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 909 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-031-485-26

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-750-590-2 x US-09-031-485-26/rev ..  
Align seq 1/1 to reverse of: US-09-031-485-26 from: 1 to: 909

92 LeuH1seuAlaAlaTyrcly 99  
|||||  
282 TTGCATTACCTGCTACTATGCG 259

seq.name: /cgn2.6/ptdata/2/1ne/5A.COMB.seq:US-08-847-429A-25

seq\_documentation\_block:  
Sequence 25, Application US/08847429A  
Patent No. 5827692  
GENERAL INFORMATION:  
APPLICANT: Tang, Liang  
APPLICANT: Blahm, E. Scott  
TITLE OF INVENTION: DIOFILARIA AND BRUGIA ANKYRIN  
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
TITLE OF INVENTION: USES THEREOF  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carol Talkington Verser, Ph.D.  
STREET: 1825 Sharp Point Drive  
CITY: Fort Collins  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80525  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: Wordperfect for Windows, Version 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/847.429A  
FILING DATE: 24-Apr-1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: HW-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 970/484-9505  
TELEFAX: 970/484-9505  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 909 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-847-429A-25

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-750-590-2 x US-08-847-429A-25 ..  
Align seq 1/1 to: US-08-847-429A-25 from: 1 to: 909

92 LeuH1seuAlaAlaTyrcly 99  
|||||  
628 TTGCATTACCTGCTACTATGCG 651

seq.name: /cgn2.6/ptdata/2/1ne/5A.COMB.seq:US-08-847-429A-26

seq\_documentation\_block:  
Sequence 26, Application US/08847429A  
Patent No. 5827692  
GENERAL INFORMATION:  
APPLICANT: Tang, Liang  
APPLICANT: Blahm, E. Scott  
TITLE OF INVENTION: DIOFILARIA AND BRUGIA ANKYRIN  
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
TITLE OF INVENTION: USES THEREOF  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carol Talkington Verser, Ph.D.  
STREET: 1825 Sharp Point Drive  
CITY: Fort Collins  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80525  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: Wordperfect for Windows, Version 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/847.429A  
FILING DATE: 24-Apr-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: HW-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 970/484-9505  
TELEFAX: 970/484-9505  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 909 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear



```

? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? TITLE OF INVENTION: US$ THERIOF
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS: CAROL TALKINGTON Verset, Ph.D.
? ADDRESS: HeKa Corporation
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Wordperfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/031,485
? FILING DATE:
? CLASSIFICATION: 530
? PRIORITY APPLICATION DATA:
? APPLICATION NUMBER: US/08/847,429
? FILING DATE: 24 APR 1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verset, Carol Talkington
? REGISTRATION NUMBER: 37,459
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1000
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..909
? US-09-031-485-22

```

```

alignment_scores:
? Quality: 8.00 Length: 8
? Gaps: 0
? Ratio: 1.000 Percent identity: 100.000

```

alignment\_block:

US-09-750-590-2 x US-09-031-485-22 ..

Align seg 1/1 to: US-09-031-485-22 from: 1 to: 911

```

92 LeuHlsteuAlaAlaIaIaTyGly 99
|||||
628 TTCGATTTCCTGCTGACTGATGCG 651

```

seq\_name: /cgn2/6/prodata/2/lna/5a\_COMB.seq:US-09-031-485-24

seq\_documentation\_block:

Sequence 22, Application US/09031485

Patent No. 5824306

GENERAL INFORMATION:

APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scott

TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verset, Ph.D.

ADDRESSER: HeKa Corporation

```

? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Wordperfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/031,485
? FILING DATE:
? CLASSIFICATION: 530
? PRIORITY APPLICATION DATA:
? APPLICATION NUMBER: US/08/847,429
? FILING DATE: 24 APR 1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verset, Carol Talkington
? REGISTRATION NUMBER: 37,459
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 24:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1000
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-031-485-24

```

```

alignment_scores:
? Quality: 8.00 Length: 8
? Gaps: 0
? Ratio: 1.000 Percent identity: 100.000

```

alignment\_block:

US-09-750-590-2 x US-09-031-485-24/rev ..

Align seg 1/1 to reverse of: US-09-031-485-24 from: 1 to: 911

```

92 LeuHlsteuAlaAlaIaIaTyGly 99
|||||
284 TTCGATTTCCTGCTGACTGATGCG 261

```

seq\_name: /cgn2/6/prodata/2/lna/5a\_COMB.seq:US-08-847-429A-22

seq\_documentation\_block:

Sequence 22, Application US/08847429A

Patent No. 5827692

GENERAL INFORMATION:

APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scott

TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verset, Ph.D.

ADDRESSER: HeKa Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0



```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/847-429A
? FILING DATE: 24-APR-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Versef, Carol Talkington
? REGISTRATION NUMBER: 37,459
? REFERENCE/DOCKET NUMBER: HM-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 911 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..909
? US-08-847-429A-22

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
? US-09-750-590-2 x US-08-847-429A-22 ..
?
? Align seq 1/1 to: US-08-847-429A-22 from: 1 to: 911
?
? 92 LeuHisIeuAlaAlaIleuTyrGly 99
? |||||
? 628 TTGCATTTCAGCTCTACATGATGTC 651

seq.name: /cgn2.6/pdata2/1na/5A.COMB.seq:us-08-847-429A-24

seq.documentation_block:
? Sequence 24, Application US/08847429A
? Patent No. 6063599
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? APPLICANT: Blahm, E. Scot
? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? TITLE OF INVENTION: USES THEREOF
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESS: Carol Talkington Versef, Ph.D.
? ADDRESS: Hask Corporation
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: Wordperfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/847-429A
? FILING DATE: 24-APR-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Versef, Carol Talkington
? REGISTRATION NUMBER: 37,459
? REFERENCE/DOCKET NUMBER: HM-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272

```

```

? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 24:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 911 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-08-847-429A-24

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
? US-09-750-590-2 x US-08-847-429A-24/rev ..
?
? Align seq 1/1 to reverse of: US-08-847-429A-24 from: 1 to: 911
?
? 92 LeuHisIeuAlaAlaIleuTyrGly 99
? |||||
? 284 TTGCATTTCAGCTCTACATGATGTC 261

seq.name: /cgn2.6/pdata2/1na/6A.COMB.seq:us-09-065-474-22

seq.documentation_block:
? Sequence 22, Application us/09065474
? Patent No. 6063599
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? APPLICANT: Blahm, E. Scot
? TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? TITLE OF INVENTION: USES THEREOF
? NUMBER OF SEQUENCES: 171
? CORRESPONDENCE ADDRESS:
? ADDRESS: Carol Talkington Versef, Ph.D.
? ADDRESS: Haska Corporation
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: Wordperfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/065-474
? FILING DATE: 24-APR-1998
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Versef, Carol Talkington
? REGISTRATION NUMBER: 37,459
? REFERENCE/DOCKET NUMBER: HM-5-C1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 911 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..909
? US-09-065-474-22

```



```

1  TITLE OF INVENTION: DIOFOLIA AND BRUGIA ANKRYN
2  TITLE OF INVENTION: PROTEIN; NUCLEOTIC ACID MOLECULES; AND
3  NUMBER OF INVENTION: US 5,718,000
4  NUMBER OF SEQUENCES: 171
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Carol Talkington Versar, Ph.D.
7  STREET: 1825 Sharp Point Drive
8  CITY: Fort Collins
9  STATE: Colorado
10 COUNTRY: USA
11 FILING DATE: 24 APR-1998
12 PUBLICATION DATE: 24 APR-1998
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/09/065,474
15 PRIORITY NUMBER: 09/065,474
16 ATTORNEY/AGENT INFORMATION:
17 NAME: Versar, Carol Talkington
18 REGISTRATION NUMBER: 37,459
19 REFERENCE/DOCKET NUMBER: HW-5-C1
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: 970/493-7272
22 TELEFAX: 970/484-9505
23 INFORMATION: 970/493-7272
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 1056 nucleotides
26 TYPE: nucleic acid
27 STRANDEDNESS: single
28 TOPOLOGY: linear
29 MOLECULE TYPE: cDNA
30 US-09-065-474-140

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x US-09-065-474-140/rev ..
Align seg 1/1 to reverse of: US-09-065-474-140 from: 1 to: 1056
92 lenhlslemlalalstyrigly 99
|||||
717 TTCACATTTCAGCTCCTAGATATGCC 694

seq_name: /cgn2/6/prodata/2/lna/5B.COMB.seq:US-08-981-729-5
seq_documentation_block:
1 Sequence 5, Application US/08981729
2 Patent No. 5,718,000
3 GENERAL INFORMATION:
4 APPLICANT:
5 TITLE OF INVENTION: A novel method to isolate mutants and to
6 NUMBER OF SEQUENCES: 10
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: floppy disk
9 OPERATING SYSTEM: IBM PC compatible
10 SOFTWARE: PC-DOS/MS-DOS
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/981,729
13 FILING DATE:
14 INFORMATION FOR SEQ ID NO: 5:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 2054 base pairs

```

```

1  TYPE: nucleic acid
2  STRANDEDNESS: double
3  TOPOLOGY: linear
4  MOLECULE TYPE: DNA (genomic)
5  HYPOTHETICAL: NO
6  ANTI-SENSE: NO
7  ORIGINAL SOURCE:
8  ORGANISM: Aspergillus tubingensis
9  PATHOSOURCE:
10 NAME/KEY: TATA, signal
11 LOCATION: 848..854
12 FEATURE:
13 NAME/KEY: exon
14 LOCATION: 950..1179
15 FEATURE:
16 NAME/KEY: intron
17 LOCATION: 1179..1228
18 FEATURE:
19 NAME/KEY: exon
20 LOCATION: 1229..1631
21 FEATURE:
22 NAME/KEY: CDS
23 LOCATION: join(950..1179, 1229..1631)
24 OTHER INFORMATION: /EC_number=3.2.1.8
25 OTHER INFORMATION: /product="1,4-beta-xylanxylohydrolase"
26 OTHER INFORMATION: /gene="xlnA"
27 OTHER INFORMATION: /standard_name="endo-xylanase"
28 FEATURE:
29 NAME/KEY: mat.peptide
30 LOCATION: 1031..1631
31 FEATURE:
32 NAME/KEY: a19-peptide
33 LOCATION: 950..1031
34 US-08-981-729-5

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x US-08-981-729-5/rev ..
Align seg 1/1 to reverse of: US-08-981-729-5 from: 1 to: 2054
939 serclnglslgsllevalthrlheu 946
|||||
1636 TCTCAGAGAGCATGTGTGACACTA 1613

seq_name: /cgn2/6/prodata/2/lna/5A.COMB.seq:US-07-842-349-1
seq_documentation_block:
1 Sequence 1, Application US/07842349
2 Patent No. 5,358,866
3 GENERAL INFORMATION:
4 APPLICANT: VAN DEN BROECK, HENRIETTE C.
5 APPLICANT: DE GRAEF, HENRIETTE C.
6 APPLICANT: HILLE R., JAN D.
7 APPLICANT: VAN OYEN J., ALBERT J.
8 APPLICANT: VISSER, JACOB
9 APPLICANT: WANDER, JACOB
10 APPLICANT: WANDER, JACOB
11 TITLE OF INVENTION: FROM FUNGAL ORIGIN
12 NUMBER OF SEQUENCES: 23
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: MORRISON & FORSTER
15 STREET: 755 Page Mill Road
16 CITY: Palo Alto
17 STATE: California
18 COUNTRY: USA
19 ZIP: 94304-1018
20 COMPUTER READABLE FORM:

```



```

? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2814 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? FEATURE: CDS
? LOCATION: 1617..2813
? PCT-US92-08328-1

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
? US-09-750-590-2 x PCT-US92-08328-1 ...
? Align seg 1/1 to: PCT-US92-08328-1 from: 1 to: 2814
?
? 1020 llaletnrlleuclnlyslapleu 1027
? |||||
? 2773 ATTTCACCTTACACAAAGATTTA 2796

seq_name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:US-09-031-485-35

seq_documentation_block:
? Sequence 35, Application US/09031485
? Patent No. 5824306
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? APPLICANT: Blum, E. Scott
? TITLE OF INVENTION: DIROPHILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? FILING DATE: 05/09/031.485
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/847,429
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 7,459
? ADDRESS/DOCK NUMBER: 148-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 35:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5235 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single

```

```

? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-031-485-35

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
? US-09-750-590-2 x US-09-031-485-35 ..
? Align seg 1/1 to: US-09-031-485-35 from: 1 to: 5235
?
? 92 leuhtleuclnlyslapleu 99
? |||||
? 1645 TTGCTTACCTTACACAAAGATTTC 1668

seq_name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:US-09-031-485-36

seq_documentation_block:
? Sequence 36, Application US/09031485
? Patent No. 5824306
? GENERAL INFORMATION:
? APPLICANT: Tang, Liang
? APPLICANT: Blum, E. Scott
? TITLE OF INVENTION: DIROPHILARIA AND BRUGIA ANKYRIN
? TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carol Talkington Verser, Ph.D.
? STREET: 1825 Sharp Point Drive
? CITY: Fort Collins
? STATE: Colorado
? COUNTRY: USA
? ZIP: 80525
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: Windows 95
? SOFTWARE: WordPerfect for Windows, Version 7.0
? CURRENT APPLICATION DATA:
? FILING DATE: 05/09/031.485
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/847,429
? FILING DATE: 24-APR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Verser, Carol Talkington
? REGISTRATION NUMBER: 7,459
? ADDRESS/DOCK NUMBER: 148-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 970/493-7272
? TELEFAX: 970/484-9505
? INFORMATION FOR SEQ ID NO: 36:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5235 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-031-485-36

alignment_scores:
? Quality: 8.00 Length: 8
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

```



STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80515  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Wordperfect for Windows, Version 7.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/065,474  
 FILING DATE: 24-APR-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verser, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/493-7272  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 35:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5235 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-09-065-474-35

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x US-09-065-474-35 ..  
 Align seg 1/1 to: US-09-065-474-35 from: 1 to: 5235

92 LeuHSLcuaAlaAlaLysTyrGly 99  
 |||||  
 1645 TTCGATTAGCTGCTAGATATGCC 1668

seq\_name: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:us-09-065-474-36

seq\_documentation\_block:  
 Sequence 36, Application US/09065474  
 Patent No. 6061599  
 GENERAL INFORMATION:  
 APPLICANT: Tang, Liang  
 APPLICANT: Blehm, E. Scott  
 TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 NUMBER OF INVENTIONS: US5 THEREOF  
 NUMBER OF SEQUENCES: 111  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Carol Talkington Verser, Ph.D.  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80525  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Wordperfect for Windows, Version 7.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/065,474  
 FILING DATE: 24-APR-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 REFERENCE/DOCKET NUMBER: HW-5-C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/493-7272  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 36:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5235 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-09-065-474-36

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x US-09-065-474-36/rev ..  
 Align seg 1/1 to reverse of: US-09-065-474-36 from: 1 to: 5235

92 LeuHSLcuaAlaAlaLysTyrGly 99  
 |||||  
 3591 TTCGATTAGCTGCTAGATATGCC 3568

seq\_name: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:us-09-031-485-32

seq\_documentation\_block:  
 Sequence 32, Application US/09031485  
 Patent No. 6243030

GENERAL INFORMATION:  
 APPLICANT: Tang, Liang  
 APPLICANT: Blehm, E. Scott  
 TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 NUMBER OF INVENTIONS: US5 THEREOF  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Carol Talkington Verser, Ph.D.  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80525  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Wordperfect for Windows, Version 7.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/031,485  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/09/847,429  
 FILING DATE: 24-APR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verser, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/493-7272  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5503 nucleotides  
 TYPE: nucleic acid







**THIS PAGE BLANK (USPTO)**





JOURNAL  
COMMENT

EST discovery in ovine  
Unpublished (2000)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smitht@mail.marc.usda.gov  
This was sequencing bases called and trimmed with phred  
3.0 (900bp) and then aligned by cross-match with the minscore 20  
and minmatch 12 options.

PCR Primers:  
FORWARD: AGGAAACAGCATGACCAT  
BACKWARD: GTTTCGACATGACAGC  
Plate: 10 row: L column: 2  
Seq primer: ATTAGGTGACATGAC.  
Location/Qualifiers

source  
1..399 sm="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 2P16"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCWV SPORT6; Site.1: XbaI; Site.2: XhoI;  
Library made from pooled tissue from testis, ovary,  
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 171 a 62 c 89 g 77 t  
ORIGIN

alignment\_scores:  
Quality: 80.00 Length: 80  
Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-750-590-2 x AM478133

Align seg 1/1 to: AM478133 from: 1 to: 399

```

1226 LeuAaNgLuLysLeuAlaAaenLeuAaNaNgLyTYrGtGluGluValCysG1 1242
|||||
158 TTAATAAGAAAATATGCCCATCTCAATAGCAATGATGAGCAAGTATGTGA 207
|||||
1242 uGluValLeuHisAlaLysLysGtGluLeuSerAlaLysAspGluLysG 1259
|||||
208 GAGGATTGATGCCAAGAAAGAGAGATGTCCTCCAGAAAGATACAGAG 257
|||||
1259 LuLeuLeuHisPheSerTLeGluGluGluLysAspGluGluLysG 1275
|||||
258 AATTCGCTCATTCAGCATTCAGACAGAAATCAAGATCAGACAGAAACGA 307
|||||
1276 CysAspLysSerLeuThrThrTLeThrGluLeuGlnAgaTgTLeGlnG1 1292
|||||
308 TCCGATTAAGTCTTACACACATCAACAGAGTTCACAGAAAGATACAG 357
|||||
1292 uSerAlaLysGlnTLeGluLysAspAsnLysTLeThr 1305
|||||
358 ATCCGCCAACAATTCAGCCGAAGATTAAGATTAAGTAACT 397

```

seq\_name: gb\_esc2:BF574189  
seq\_documentation\_block: 863 bp mRNA  
DEFINITION 602131348P1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4271074 5',  
mRNA sequence.  
ACCESSION BF574189  
VERSION BF574189.1 GI:11647901  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
BANKIT:10, EMBL:10, Chordata: Craniata, Vertebrata: Euteleostomi;  
Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.

REFERENCE  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgsps-re@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Prepared by: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LULM)  
DNA Sequencing by: Invitae Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LULM at:  
<http://image.llnl.gov>  
Plate: LUCM1082 row: 0 column: 11  
High quality sequence stop: 559.  
Location/Qualifiers

source  
1..863 sm="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:4271074"  
/clone\_lib="NIH\_MGC\_81"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: muscle (skeletal); Vector: pNR-11B  
(Clontech); Site.1: SfiI (ggcgccgcggccg); Site.2: SfiI  
(ggcgccgcggccg); 5' and 3' adaptors were used in cloning  
as follows: 5' adaptor sequence: 5'-CACGCCATATGACC-3'  
and 3' adaptor sequence: 5'-ATTCGACGCGCGCGCGACATG-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size  
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained  
inserts by PCR. This library was enriched for full-length  
clones and was constructed by Clontech Laboratories (Palo  
Alto, CA)."

BASE COUNT 324 a 138 c 233 g 168 t  
ORIGIN

alignment\_scores:  
Quality: 64.00 Length: 64  
Ratio: 1.000 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-750-590-2 x BF574189

Align seg 1/1 to: BF574189 from: 1 to: 863

```

451 GtLeuAlaHisLysValAlaGluTLeCysLysAlaLeuAlaLeuGluLysG1 467
|||||
1 GAACTGCGACACAAAGTGGCAATGCAAAACCTTTCAGTATGAAATGTA 50
|||||
467 uATGValLysGluAspSerAspGluGlnTLeLysGlnLeuGluAspAla 484
|||||
51 AAGGCTCAAGAGAGATTCATCAACACATGACATGACATGAGAAAGTCA 100
|||||
484 GtLysAspArgGlnLysArgMetTLeGluLeuGluLysGluValLysGln 500
|||||
101 TAAAGAGTGTCCAGAAAGAGATGATGATGATCAAGGTTAAAGTTAAACA 150
|||||
501 MetGlnThrHisPheLeuAlaLeuLysGluHisLysLeuThrSer 514
|||||
151 ATGCAAGCCCATTTCTTGCCCTTAAAGAACACTTAACAAAGCT 192

```

seq\_name: gb\_esc1:BE350494  
seq\_documentation\_block: 468 bp mRNA  
LOCUS BE350494  
DEFINITION h14612.x1 NCI-CGAP\_K1d3 Homo sapiens cDNA clone IMAGE:3146734 3',  
similar to TR:028282\_Q28282 CAVS PROTEIN. mRNA sequence.  
ACCESSION BE350494  
VERSION BE350494.1 GI:9262347  
KEYWORDS EST.  
SOURCE human.



JOURNAL  
COMMENT: Unpublished (1999)  
Contact: Robert Streusberg, Ph. D.  
Email: craps@femail.nlh.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MCC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILML at:  
http://imgcni.nih.gov  
Library: 190293  
Library Size: P column: 05  
High quality sequence stop: 690.  
Location/Qualifiers

FEATURES  
source  
1. 883  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4472692"  
/clone\_1b="NH-MCC-90"  
/tissue\_type="adenoarcinoma cell line"  
/note="CDNA library (pharmaceutical)"  
/note="Organ. Livers; Vector: PCMV-SPOB6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally; oligo-dr primed;  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MCC Library."  
BASE COUNT 305 a 159 c 207 g 212 t  
ORIGIN

alignment\_scores:  
Quality: 55.00 Length: 55  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-750-590-2 x BG253578  
Align seg 1/1 to: BG253578 from: 1 to: 883

277 GATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 237  
|||||  
188 CAGAGGAGCATCAAAATTTGAGATTGGAATTAAGCAATTT 237  
|||||  
293 ATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 310  
|||||  
238 GAGAGGAGCATCAAAATTTGAGATTGGAATTAAGCAATTT 287  
|||||  
310 TAAATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 326  
|||||  
288 AATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 337  
|||||  
327 AATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 337  
|||||  
338 GATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 352  
seq\_name: gb\_est1:AA329627  
seq\_documentation\_block:  
LOCUS AA329627 373 bp mRNA EST 20-APR-1997  
DEFINITION EST329627 Embryo, 12 week II Homo sapiens CDNA 5' and similar to  
similar to myosin heavy chain family, mRNA sequence.  
ACCESSION AA329627  
KEYWORDS EST 329627.1 GI:1981890  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 373)  
Adams,M.D., Kertlage,A.R., Eletschmann,R.D., Fuldner,R.A., Bult,  
C.J., Lee,M.H., Kirkness,E., McIninch,K.G., Roach,J.C., Clayton,R.A.,  
Blanton,J., Blanton,J., Blanton,J., Blanton,J., Blanton,J., Blanton,J.,  
Clive,T.R., Cotton,M.D., Earle-Hughes,E., Fine,L.D., Fitzgerald

JOURNAL  
COMMENT: Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
MEDLINE  
96026280  
OBJECT: Kertlage, AR  
BIOINFORMATICS  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arker@leig.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the t10k human gene  
index (http://www.ncbi.nlm.nih.gov/ncbi/ncbi.html)  
Seq primer: M3 Reverse  
Location/Qualifiers

FEATURES  
source  
1. 373  
/organism="Homo sapiens"  
/db\_xref="ATCC (host):131302"  
/db\_xref="taxon:9606"  
/clone\_1b="Embryo, 12 week II"  
/note="Stage: Embryo, 12 week II"  
/note="Vector: plasmid; SK-; Site.1: EcoRI; Site.2:  
XhoI"  
BASE COUNT 149 a 58 c 90 g 74 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 53.00 Length: 53  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-750-590-2 x AA329627  
Align seg 1/1 to: AA329627 from: 1 to: 373

440 GATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 456  
|||||  
22 TCGAGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 71  
|||||  
456 TACGAGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 121  
|||||  
72 GCGAGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 121  
|||||  
473 GATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 489  
|||||  
122 GATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 171  
|||||  
490 ATGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 492  
|||||  
172 AGGAGTAAAGTAAATTTGAGATTGGAATTAAGCAATTT 180  
seq\_name: gb\_est1:HA5100  
seq\_documentation\_block:  
LOCUS HA5100 485 bp mRNA EST 31-JUL-1995  
DEFINITION y066405\_r1 Soares Breast 3JNH8st Homo sapiens cDNA clone  
IMAGE:182864 5', mRNA sequence.

ACCESSION H45100  
 VERSION H45100.1 GI:921152  
 EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Ellstrand, K., Hawley, M., Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, C., Marr, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Mohlman, P. and Wilson, R.  
 TITLE The Marsh-Merck EST Project  
 JOURNAL Proc Natl Acad Sci USA  
 COMMENT Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Insert Size: 645  
 High quality sequence stops: 338  
 SOURCE: IMAGE Consortium, LLMU  
 This is a publicly available, free through LLMU; contact the IMAGE Consortium (info@image.llnu) for further information.  
 Insert Length: 645 Std Error: 0.00  
 Seq primer: M13rpl  
 High quality sequence stop: 338.  
 Location/Qualifiers  
 1. 485  
 /organism="Homo sapiens"  
 /db\_xref="Gene:381804"  
 /db\_xref="Gene:381805"  
 /clone="IMAGE:182864"  
 /clone.lib="Soares breast 3mbHst"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: breast; Vector: p773D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was ligated into a Not I oligo(dT) primer 15' TCTTACCACTGATGATCCTTAAACAACATCAACGCTTACAAAGAAAT 130 double-stranded cDNA was ligated to Eco RI (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fallina Bomalido."  
 BASE COUNT 166 a 95 c 115 g 105 t 4 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 53.00 Length: 53  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-750-590-2 x H45100 ..  
 Align seg 1/1 to: H45100 from: 1 to: 485  
 1274 GUAUGCysApPySSeLeuThrTrIleThnGluLeuGlnAATGATG 1290  
 |||||||  
 81 GAGCAGATGATGATGATCCTTAAACAACATCAACGCTTACAAAGAAAT 130  
 |||||||  
 1290 GAGGluSeRAlaLySGlnIleGlnAlaLyAspAsnLyIleThnGlu 1307  
 |||||||  
 131 ACAAGAACTCTGTAACCAATTCAGAACCAAGATTAATAGATACGAAAC 160  
 |||||||  
 1307 euleuAsnAspValGluArgLeuLySGlnAlaLeuAsnGlyLeuSerGln 1323  
 |||||||  
 181 TECTTAATGATGTCGAACAAATTAAACAGGACTCAATGCCCTTCCCA 210

1324 LeuThrTrYr 1326  
 |||||||  
 231 CTCACCTAC 239  
 seq\_name: gb\_est2:BC540419  
 seq\_documentation\_block: 609 bp mRNA  
 DEFINITION BC540419.1 NIH\_MGC.77 Homo sapiens cDNA IMAGE:4693423 5', 3' UTR sequence.  
 ACCESION BC540419  
 VERSION BC540419.1 GI:13532652  
 EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE  
 AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: c9apbs-f@mail.nih.gov  
 Tissue Procurement: CLOUTECH Laboratories, Inc.  
 cDNA Library Preparation: CLOUTECH Laboratories, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLMU)  
 DNA Sequencing: NIH  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLMU at: http://image.llnu.gov  
 Plate: LCM1515 row: m column: 08  
 High quality sequence stop: 564.  
 Location/Qualifiers  
 1. 609  
 /organism="Homo sapiens"  
 /db\_xref="Gene:381804"  
 /db\_xref="Gene:381805"  
 /clone="IMAGE:4693423"  
 /clone.lib="NIH-MGC-77"  
 /lab\_host="DH10B (71 phage-resistant)"  
 /note="Organ: lung; Vector: pNR-LIB (Clontech); Site.1: SfiI (ggccgctggcgc); Site.2: SfiI (ggccataggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTTAGGCC-3' and 3' adaptor sequence: 5'-GAGCAGATGATGATCCTTAAACAACATCAACGCTTACAAAGAAAT 134  
 c. ori G and N (5'-4-0 kb). 12/15 colonies were picked and by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."  
 BASE COUNT 203 a 134 c 131 g 141 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 53.00 Length: 53  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-750-590-2 x BC540419 ..  
 Align seg 1/1 to: BC540419 from: 1 to: 609  
 1274 GUAUGCysApPySSeLeuThrTrIleThnGluLeuGlnAATGATG 1290  
 |||||||  
 81 GAGCAGATGATGATCCTTAAACAACATCAACGCTTACAAAGAAAT 134  
 |||||||  
 1290 GAGGluSeRAlaLySGlnIleGlnAlaLyAspAsnLyIleThnGlu 1307  
 |||||||  
 135 ACAAGAACTCTGTAACCAATTCAGAACCAAGATTAATAGATACGAAAC 184  
 |||||||  
 1307 euleuAsnAspValGluArgLeuLySGlnAlaLeuAsnGlyLeuSerGln 1323







## alignment\_scores:

Quality: 51.00 Length: 51  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-750-590-2 x BF416703/rev

Align seg 1/1 to reverse of: BF416703 from: 1 to: 336

```

1349 LeungGingInGinLeuAlaAspArgGlnHisGlnGluVal11Leu1 1365
1350 |||||||
285 CTGCACGACGACGCTGCTGATCTGCAGACACGACGACGATGTCGC 236
1365 attTyrArGThrHisLeuSerAlaAlaGlnGlyHisMetAspGluA 1382
1366 |||||||
235 TATCTACCGACGACACCTCTGACGCGACGAGGTCACATGCGAAG 186
1382 spvAlaGlnAlaAlaLeuGlnGlnIleGlnImeArGArgIndYLeuVal 1398
1383 |||||||
185 ATGTGCAGCGACGCTTCCTGTCAGATCATCATCATCGACACGACGCTCG 136
1399 Cys 1399
|||
135 TTC 133

```

seq\_name: gb-est1:AA684631

## seq\_documentation\_block:

LOCUS AA684631 337 bp mRNA EST 03-APR-1998  
DEFINITION EST104976 Rat PC-12 cells, untreated Rattus sp. cDNA clone RC4E51  
3' end, mRNA sequence.  
ACCESSION AA684631  
VERSION AA684631.1 GI:2671229  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus norvegicus  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;

## REFERENCE

## AUTHORS

Lee, N.H., Melnick, R.G., Kirness, E.F., Earle-Hughes, J.A., Fuldner  
, R.A., Mammas, S., Glodex, A., Gocayne, J.D., Adams, M.D., Kerlavage  
, A.R., Fraser, C.M. and Venter, J.C.  
Comparative expressed-sequence-tag analysis of differential gene  
expression profiles in PC-12 cells before and after nerve growth  
factor treatment. *Proc Natl. Acad. Sci. U.S.A.* 92, 8033-8037 (1995)

## TITLE

Proc Natl. Acad. Sci. U.S.A. 92, 8033-8037 (1995)  
95396786

## JOURNAL

## MEDLINE

Contact: Lee, NH  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@igmc.org  
Seq primer: M1321

## FEATURES

## source

1. 337  
/organism="Rattus sp."  
/db\_xref="ATCC:(inhost):2000223"  
/db\_xref="taxon:10116"

/clone="RC4E51"  
/clone="lib-Rat-PC-12 cells, untreated"

/clone="lib-Rat-PC-12 cells, untreated"  
/clone="lib-Rat-PC-12 cells, untreated"

/clone="lib-Rat-PC-12 cells, untreated"  
/clone="lib-Rat-PC-12 cells, untreated"

/clone="lib-Rat-PC-12 cells, untreated"  
/clone="lib-Rat-PC-12 cells, untreated"

/clone="lib-Rat-PC-12 cells, untreated"  
/clone="lib-Rat-PC-12 cells, untreated"

/clone="lib-Rat-PC-12 cells, untreated"  
/clone="lib-Rat-PC-12 cells, untreated"

BASE COUNT 73 a 93 c 94 g 76 t 1 others

## alignment\_scores:

Quality: 51.00 Length: 51  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-750-590-2 x AA684631/rev

Align seg 1/1 to reverse of: AA684631 from: 1 to: 337

```

1349 LeungGingInGinLeuAlaAspArgGlnHisGlnGluVal11Leu1 1365
1350 |||||||
271 CTGCACGACGACGCTGCTGATCTGCAGACACGACGACGATGTCGC 222
1365 attTyrArGThrHisLeuSerAlaAlaGlnGlyHisMetAspGluA 1382
1366 |||||||
221 TATCTACCGACGACACCTCTGACGCGACGAGGTCACATGCGAAG 172
1382 spvAlaGlnAlaAlaLeuGlnGlnIleGlnImeArGArgIndYLeuVal 1398
1383 |||||||
171 ATGTGCAGCGACGCTTCCTGTCAGATCATCATCATCGACACGACGCTCG 122
1399 Cys 1399
|||
121 TTC 119

```

seq\_name: gb-est2:BI276339

## seq\_documentation\_block:

LOCUS BI276339 349 bp mRNA EST 19-JUL-2001  
DEFINITION UI-R-CMD-BWP-a-05-0-01.s1 UI-R-CMD Rattus norvegicus cDNA clone  
UI-R-CMD-BWP-a-05-0-01 3', mRNA sequence.  
ACCESSION BI276339  
VERSION BI276339.1 GI:14921133  
KEYWORDS EST.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;

## REFERENCE

## AUTHORS

Bonaldi, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
9704477

## TITLE

Proc Natl. Acad. Sci. U.S.A. 92, 8033-8037 (1995)  
9704477

## JOURNAL

## MEDLINE

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu

## FEATURES

## source

1. 349  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

/clone="lib-R-CMD-BWP-a-05-0-01"  
/clone="lib-R-CMD-BWP-a-05-0-01"

```

/dev stage="ADULT"
/ncbi="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The U-R-CW0
library is a normalized library constructed from the
following rat source tissues: embryonic day 19, embryonic
day 21, adult day 1, adult day 12, adult day 75, adult day
200. For details of the library from which this clone
was derived, please visit our web site at
http://www.igmm.uio.no. The subtraction has been previously
described in (Bonaldi, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG: LIB-UI-R-CW0
TAG-TISSUE: rat, aorta pool
TAG-500-CYTACGATC*
ORIGIN
69 a 34 c 30 g 96 t

alignment_scores:
Quality: 51.00 Length: 51
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x B1276339/rev

Align seq 1/1 to reverse of: B1276339 from: 1 to: 349

1349 leuglncinglnleualaaapalaasparglnhslglncluvallleal 1365
|||||
261 ctgcacgacacacgtctctctcctcctcctcctcctcctcctcctcctc 232
1365 attctyargthrlslaueluserialaaglnclgylslmetcaspdla 1382
|||||
231 tatctaccggacacacgtctcctcctcctcctcctcctcctcctcctc 182
1382 spvalglnalalaleuenglntlletlglnmclatrglnclgylueual 1398
|||||
181 atgtcagacacacgtctctcctcctcctcctcctcctcctcctcctc 132
1399 Cys 1399
|||
131 TGC 129

seq_name: gb_cst1:A1501930
seq_documentation_block:
LOCUS A1501930 353 bp mRNA EST 03-JUL-1999
DEFINITION UI-R-C0-1V-b-05-0-UI-g1 UI-R-C0 Rattus norvegicus cDNA clone
ACCESSION UI-R-C0-1V-b-05-0-UI_3', mRNA sequence.
VERSION A1501930
KEYWORDS EST.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 353)
BONALDI,M.F., LENNON,G. and SOARES,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL
MEDLINE
COMMENT
Cytosine Res. 6 (9), 791-806 (1996)
PROGRAM FOR RATTUS NORVEGICUS
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@uiowa.edu
The sequence contained an oligo-dT track that was present in the

```

```

oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The oligo-dT track served to bridge the gap between the NotI site
and the oligo-dT track. The cDNA between the NotI site and the
normalized 18 day embryo library cDNA library preparation. M.B.
Soares Lab Clone Distribution: clones will be available through
Research Genetics (www.resgen.com) This clone is also available
through the I.M.A.G.E. Consortium at LNC (info@imgc.llnl.gov).
IMAGE ID=B1276339 The following repetitive elements were found in
this cDNA sequence: 1-39, xMT-richlow_complexity
POLY-A-NO.
location/Qualifiers
1. 353
/orrganism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C0-1V-b-05-0-UI"
/organism="Rattus norvegicus"
/dev stage="Adult"
/ncbi="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The U-R-C0
library is a subtracted library derived from the U-R-A1
and U-R-E1 libraries. The U-R-A1 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, and muscle. The U-R-E1
library consisted of a mixture of individually tagged
normalized libraries constructed from 8, 12 and 18 day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which
allows identification of the library of origin of a clone
within the mixture. The subtracted library (U-R-C0) was
constructed as follows: PCR amplified cDNA inserts from a
pool of 200 embryos and 200 adult clones from which 3 ESTs had
been derived was used as a library. The library was then
been derived from a pool of 200 embryos and 200 adult clones
the pooled U-R-A1 and U-R-E1 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the U-R-C0
library. This procedure has been previously described
(Bonaldi, Lennon and Soares, Genome Research 6: 791-806,
1996).

BASE COUNT
72 a 95 c 90 g 1 othrs
ORIGIN

alignment_scores:
Quality: 51.00 Length: 51
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-750-590-2 x A1501930/rev

Align seq 1/1 to reverse of: A1501930 from: 1 to: 353

1349 leuglncinglnleualaaapalaasparglnhslglncluvallleal 1365
|||||
261 ctgcacgacacacgtctctcctcctcctcctcctcctcctcctcctc 235
1365 attctyargthrlslaueluserialaaglnclgylslmetcaspdla 1382
|||||
234 tatctaccggacacacgtctcctcctcctcctcctcctcctcctcctc 185
1382 spvalglnalalaleuenglntlletlglnmclatrglnclgylueual 1398
|||||
184 atgtcagacacacgtctctcctcctcctcctcctcctcctcctcctc 135
1399 Cys 1399

```

```

111
134 TCC 132
seq_name: gb_est1:A1764178
seq_documentation_block: 368 bp mRNA
LOCUS A1764178
DEFINITION UI-R-Y0-act-g-05-0-UI s1 UI-R-Y0 Rattus norvegicus cDNA clone
ACCESSION A1764178
VERSION A1764178.1 GI:5210113
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Rattus; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
1 (bases 1 to 368)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoreas@iue.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a homafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalised pool of track served to verify it as a clone from the
normalised pool of track served to verify it as a clone from the
clone distribution: clones will be available through research
Genetics (www.resgen.com) The following repetitive elements were
found in this cDNA sequence: 1-23, >AT_richlow_complexity
Seq primer: M13 Forward
POLYA-Yes.
FEATURES
Source
Location/Qualifiers
1..368
/organism="Rattus norvegicus"
/organism_synonym="Norway rat"
/db_xref="taxon:10116"
/clone="UI-R-Y0-act-g-05-0-UI"
/clone_1lb="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/vector="p773D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
oligonucleotide pool of all previous libraries (UI-R-A0, UI-R-B0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
which 3' end of the cDNA inserts from previous library clones from
which 3' end of the cDNA inserts from previous library clones from
hybridization with the normalized whole genome library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(1996).
Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
TAG_L1b-UI-R-Y0

```

```

TAG_TISSUE=EYE
TAG_SEQ-CATG"
BASE COUNT 80 a 98 c 93 g 97 t
ORIGIN
alignment_scores:
Quality: 51.00 Length: 51
Percent Similarity: 100.000 Gaps: 0
Percent Identity: 100.000
alignment_block:
US-09-750-590-2 x A1764178/rev ...
Align seg 1/1 to reverse of: A1764178 from: 1 to: 368
1349 LengthGInGInLeuAlaAspAlaAspAsgInHISGInGInValAlaIleA1 1365
|||||
298 CTGGAGACAGCAGCTGAGCTGATGCTGACAGCAGCAGCAGAGTATCTCC 249
|||||
1365 ATCTGTATGTHHSLLeuLeuSer:AlaAlaGInGInGInHISMetAlaSpG1A 1382
|||||
248 TATCTACCGGACAGCAGCTCTCATGCTCCAGCCAGGCTCATGTCATGAAAG 199
|||||
1382 SPVALGInGInAlaAlaLeuLeuGInGInIleIleGInMetAArgGInGInLeuA1 1398
|||||
198 ATGTGACGACGACGCTGTGCTGCTGACATCATACAGATGCGACAGGCTCTGTC 149
|||||
1399 Cys 1399
148 TCC 146
seq_name: gb_est1:A1764036
seq_documentation_block: 370 bp mRNA
LOCUS A1764036
DEFINITION UI-R-Y0-act-g-03-0-UI s1 UI-R-Y0 Rattus norvegicus cDNA clone
ACCESSION A1764036
VERSION A1764036.1 GI:5209971
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Rattus; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
1 (bases 1 to 370)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoreas@iue.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a homafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalised pool of track served to verify it as a clone from the
clone distribution: clones will be available through research
Genetics (www.resgen.com) The following repetitive elements were
found in this cDNA sequence: 1-23, >AT_richlow_complexity
Seq primer: M13 Forward
POLYA-Yes.
FEATURES
Source
Location/Qualifiers
1..370
/organism="Rattus norvegicus"

```



[illegible]

```

1399 Cys 1399
111
98 TOC 96

seq_name: gb_estr1:BF42851

seq document: block:
LOCUS       BF42851               441 bp      mRNA
DEFINITION  U1-M-GDP-bn1-h-01-0-01_s1 NIH-BMAP-Ret4_S2 Mus musculus cDNA clone
ACCESSION   BF42851
VERSION     BF42851
KEYWORDS    U1-M-GDP-bn1-h-01-0-01 3', mRNA sequence.
SOURCE      Mus musculus
ORGANISM    Mus musculus
TAXID       10116
EST         BF42851.1 GI:11532074
COMMENT     house mouse,
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 441)
AUTHORS     Ronald M.F., Lemmon,G., and Soares M.B.
TITLE       Normalization and subcloning: two approaches to facilitate gene
           discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
COMMENT     Genbank accession: U13997
           Contact: Chlo. H
           National Institute of Mental Health
           6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
           20892-9643, USA
           Tel: 301 443 1706
           Fax: 301 443 9890
           Email: mestrnall@nih.gov
           The sequence contained an oligo-dT track that was present in the
           original submission. It was used to confirm the synthesis of first
           strand cDNA and therefore this may represent a homofide poly A
           tail. The sequence tag present in the cDNA between the NotI site
           and the oligo-dT track served to identify it as a clone from the
           embryonic retina tissue cDNA library Preparation: M.B. Soares Lab
           Clone distribution: Researchers may obtain BMAP cDNA clones from
           RESSEARCH GENETICS. It should be noted that Bento Soares is
           generating a small number of additional specialized non-redundant
           arrays of BMAP cDNA whose availability will be considered under
           special conditions. Collaborative arrangements
           Seq primer: M13 forward
           POLYA=yes.

FEATURES             Location/Qualifiers
     source            1..441
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /cmap="cdna"
                     /cmap_syn="h-01-0-01"
                     /clone_id="NIH-BMAP-Ret4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pTR3D-Pac (Pharmacia) with a modified
                     poly(I) linker. Site.1: Not I; Site.2: Eco RI; The
                     NIH-BMAP-Ret4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived, please
                     visit the following URL: http://www.geni.slow.edu.
                     TAG: LIB=NIH-BMAP-Ret4_S2
                     TAG: TISSUE=embryonic-retina
                     TAG: SEQ=CTCTAGACACA"
     BASE COUNT       93 a      111 g      126 t
     ORIGIN            111 c      111 g

Alignment scores:
Ratio: 1.000      Length: 51
Percent Similarity: 100.000      Gaps: 0
Percent Identity: 100.000

Alignment block:
005-09-750-590-2 x BF42851/rev

```

Align seg 1/1 to reverse of: BF462851 from: 1 to: 441

1349 LeungJingJinLeuAlaAspAlaAspArgGlnHisGlnGluValIleAla 1365  
 |||||||  
 288 CTCACGACGACGACGCTGCTCTGATCTGACGACGACCAAGGACTTATGCG 239  
 1365 aileTyrArgThrHisLeuLeuSerAlaAlaGlnGlyHisMetAspGln 1382  
 |||||||  
 238 TATCTACCGGACACGCTCTCTCTGAGTCCGCGACAGGCCCTCATGATGAG 189  
 |||||||  
 1382 spvAlGlnAlaAlaLeuLeuGlnIleIleIleMetLeuArgGlnGlyLeuVal 1388  
 |||||||  
 188 ATGTCTACGACGACGCTCTCTGATCTGATCATACGATGCGCAAGGCTTGTG 139  
 |||||||  
 1399 Cys 1399  
 |||  
 138 TCC 136

seq\_name: gb\_est1:AA153070

seq\_documentation\_block: 443 bp mRNA EST 11-FEB-1997  
 LOCUS AA153070  
 DEFINITION Mus musculus mouse heart (#937316) Mus musculus cDNA clone  
 accession: AF153070  
 version: AA153070  
 keywords: AA153070.1 GI:1724720  
 EST.  
 source: house mouse.  
 organism: Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;  
 1 (bases 1 to 443)  
 M. Mus musculus L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
 Giesel S., Kucaba T., Leach M., Le M., Le M., J. Koutis H.,  
 Schellenberg K., Stepien M., Tan F., Underwood J., Koutis H.,  
 Theisberg B., Wyle T., Lennon G., Soares B., Wilson R. and  
 Waterston R.  
 The WashU-HMI Mouse EST Project  
 unpublished (1996)  
 contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1810  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through DML: contact the  
 IMAGE Consortium (info@image.jml.gov) for further information.  
 NC1:369521  
 Seq primer: -28m13 rev1 ET from Amerham  
 High quality sequence EDP: 374.  
 Location/Qualifiers  
 1..443

## FEATURES

source

/organism="Mus musculus"  
 /strain="NIH/Swiss"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:604189"  
 /clone\_1lb="Strategene mouse heart (#937316)"  
 /seq\_strategy="heart"  
 /dev\_stage="13 day embryos"  
 /note="Organ: heart; Vector: pBluescript SK-; Site:1;  
 EcoRI; Site:2; XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.  
 Average insert size: 1.0 kb. Uni-ZAP XR Vector: -5'  
 adaptor sequence: 5' GAATTCGCGCACG 3' -3' adaptor  
 sequence: 5' CTCGACTTTTCTTTTCTTTTCTTTT 3'  
 87 a 119 c 119 g 118 t

## alignment\_scores:

Quality: 51.00 Length: 51  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

us-09-750-590-2 x AA153070/rev ..

Align seg 1/1 to reverse of: AA153070 from: 1 to: 443

1349 LeungJingJinLeuAlaAspAlaAspArgGlnHisGlnGluValIleAla 1365  
 |||||||  
 253 CTCACGACGACGCTGCTGATCTGACGACGACCAAGGACTTATGCG 204  
 |||||||  
 1365 aileTyrArgThrHisLeuLeuSerAlaAlaGlnGlyHisMetAspGln 1382  
 |||||||  
 203 TATCTACCGGACACGCTCTCTGATCTGATCATACGATGCGCAAGGCTTGTG 154  
 |||||||  
 1382 spvAlGlnAlaAlaLeuLeuGlnIleIleIleMetLeuArgGlnGlyLeuVal 1398  
 |||||||  
 153 ATCTCAGGCGACGCTCTCTGATCTGATCATACGATGCGCAAGGCTTGTG 104  
 |||  
 1399 Cys 1399  
 |||  
 103 TCC 101

seq\_name: gb\_est1:AF1576759

seq\_documentation\_block: 449 bp mRNA EST 05-APR-1999  
 LOCUS AF1576759  
 DEFINITION UI-R-Y0-V1-d-04-0-01.s1 UI-R-Y0 Rattus norvegicus cDNA clone  
 UI-R-Y0-V1-d-04-0-01-3', mRNA sequence.  
 accession: AF1576759  
 version: AF1576759.1 GI:4561135  
 keywords: AF1576759.1 GI:4561135  
 EST.  
 source: Norway rat.  
 organism: Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;  
 1 (bases 1 to 449)  
 Rattus norvegicus  
 Bonaldo M.F., Lennon G. and Soares M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 431 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel.: 319 335 8250  
 Fax: 319 335 9555  
 Email: moares@blue.wesg.uiowa.edu  
 This sequence contained an oligo-dT track that was present in the  
 original submission. The sequence was re-analyzed to determine if the  
 strand cDNA and therefore this sequence was a cDNA between the NCBI site  
 and the oligo-dT track served to verify it as a clone from the  
 normalized Eye library cDNA library Preparation: M.B. Soares Lab  
 Clone distribution: clones will be available through Research  
 Genetics (www.resgen.com) The following repetitive elements were  
 found in this cDNA sequence: 1-23, >AT-richtlow-complexity  
 Seq primer: A13 forward.

## FEATURES

source

/organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-Y0-V1-d-04-0-01"  
 /clone\_1lb="UI-R-Y0"  
 /dev\_stage="adult"  
 /note="Vector: pT730-pac (Pharmacia) with a modified  
 polylinker. Site:1; Not I; Site:4; Eco RI; The UI-R-Y0



library is a subtracted library derived from an individually prepared rat brain library. The subtracted library was prepared for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from the subtracted library were ligated into a BamHI-SmaI hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Lemon, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 96 a 115 c 112 g 126 t

ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 51  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x A1576759/rev

Align seg 1/1 to reverse of: A1576759 from: 1 to: 449

1349 LeuGInGInGInLeuAlaAPaAsPaRGInHISGInGInUValIleAl 1365  
|||||  
298 CTGCAGACGACGCTGCTGATCTGCTGACACACGACACGACGATCTGC 249  
1365 ATlETyPArPThHISleuLeuSerAlAlAlGInGInGInHISMeTArPQuA 1382  
|||||  
248 TATCTACGGGACACACTCTCTACGTCCGACAGGGGTCAATGATGATGAAG 199  
1382 sPvAlGInAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1398  
|||||  
198 ATGTGACGACGACGCTGCTGCTGCTGACATCATCATGATGACAAAGGCTCTGTG 149  
1399 Cys 1399  
|||  
148 TTC 146

seq\_name: gb\_cet2:BE984856

seq\_documentation\_block:

LOCUS BE984856 465 bp mRNA EST  
DEFINITION UI-M-CDOP-bgn-e-01-0-UI-37 UI-R-BM4-Ret4-S2 Mus musculus cDNA clone  
ACCESSION BE984856  
VERSION BE984856.1 GI:10657506  
KEYWORDS EST.

SOURCE  
ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Baltimore F. Lennon G. and Soares M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEMLINE 97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20852-9643, USA

Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mestr@mail.nih.gov  
The sequence contained an oligo-dr track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dr track served to identify the clone as a mouse  
embryonic fibroblast library. The library was prepared by M.R. Soares Lab  
Research Genetics. It should be noted that Benio Soares is  
generating a small number of additional specialized non-redundant  
arrays of BM4 cDNAs whose availability will be considered under  
appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA-Yes

FEATURES  
source  
location/Qualifiers

/organism="Mus musculus"  
/strain="C57Bl/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CDOP-bgn-e-01-0-UI-37"  
/clone\_id="N1R\_BM4-Ret4-S2"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT775-Pac (Pharmacia) with a modified  
polyA tail. The library was prepared by M.R. Soares Lab  
Research Genetics. It should be noted that Benio Soares is  
generating a small number of additional specialized non-redundant  
arrays of BM4 cDNAs whose availability will be considered under  
appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA-Yes  
TAG-TISSUE=embryonic-retina  
TAG-LIB=N1R\_BM4-Ret4-S2  
TAG-SEQ=CTGCTGATGCA 131 t

BASE COUNT 94 a 123 c 117 g

ORIGIN

alignment\_scores:  
Quality: 51.00 Length: 51  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-750-590-2 x BE984856/rev

Align seg 1/1 to reverse of: BE984856 from: 1 to: 465

1349 LeuGInGInGInLeuAlaAPaAsPaRGInHISGInGInUValIleAl 1365  
|||||  
287 CTGCAGACGACGCTGCTGATCTGCTGACACACGACACGACGATCTGC 238  
1365 ATlETyPArPThHISleuLeuSerAlAlAlGInGInGInHISMeTArPQuA 1382  
|||||  
227 TATCTACGGGACACACTCTCTACGTCCGACAGGGGTCAATGATGATGAAG 188  
1382 sPvAlGInAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1398  
|||||  
187 ATGTGACGACGACGCTGCTGCTGCTGACATCATCATGATGACAAAGGCTCTGTG 138  
1399 Cys 1399  
|||  
137 TGC 135

seq\_name: gb\_cet1:BE109950

seq\_documentation\_block:

LOCUS BE109950 485 bp mRNA EST  
DEFINITION UI-R-BJ1-aum-c-06-0-UI-31 UI-R-BJ1 Rattus norvegicus cDNA clone  
ACCESSION BE109950  
VERSION BE109950.1 GI:8502055  
KEYWORDS EST.





```
Location/Qualifiers
1..604
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U1-MC89-bqy-e-08-0-01"
/clone_id="M18.BMP.Ret4.S2"
/lab_note="DH08 (Life Technologies)"
/molecule="pT7303-Pac (Pharmacia)"
/polymer: Site_1: Not 1; Site_2: Eco RI; The
```

NHL BMP-Relt-S2 library is a subtracted library,  
ultimately derived from mouse retina tissue libraries at  
various stages of development. For a detailed description  
visit the library web page from which this clone was derived, please  
[http://www.nhl.nih.gov/research/relt\\_s2](http://www.nhl.nih.gov/research/relt_s2)  
TAC: LIB-NIH\_BMP-Relt\_S2  
TAC: TISSUE-adult-retina  
TAC\_Seq-CGCGCCGCAC\*

BASE COUNT      114 a    162 c    153 g    174 t    1 others

---

alignment\_scores:

|                     | Quality:          | 51.00   | Length: | 51      |
|---------------------|-------------------|---------|---------|---------|
| Percent Similarity: | Ratio:            | 1.000   | Gaps:   | 0       |
|                     | Percent Identity: | 100.000 |         | 100.000 |

alignment\_block:  
US-09/-750-550-2 x BF466442/rev ..

Align seg 1/1 to reverse of: BF466442 from: 1 to: 604

```

1349 LeunglncglLlemlAlaAspArglnHnslagLuvallllgaal 1365
|||||
288 CTCGACAGCAGCTGGCTATGTCGACAGCACCAACAAGCTATTGCC 239
|||||
1365 attgtyatgttttlstleuleusecrlalaaglnclnglyshmetaspqlua 1382
|||||
238 TATCTTCGCGACAAACCTCTCATGCTGACAGGCGACAGCTGATGAAG 189
|||||
1382 spvalglmlaialaleuandililelleldlmelarglnclnyluaval 1398
|||||
188 ATGTGACAGCGACCCCTACTGCAATCATACATGACGACCAAGCCTTGTS 139
|||||
1399 Gcg 1399
|||
138 Tgc 136

```

seq\_name: gb.esl2:BF722070

---

seq\_documentation\_block: 634 bp mRNA EST 03-JAN-2001

LOCUS BF722070 mblacg9701 Sources\_MBLA-BDarchival\_arch Mus musculus CDNA Clone

DEFINITION mblacg9701B.S similar to TR-09j3r5 09j3r5 HYPOTHETICAL 39.8 KD

PROTEIN , mRNA sequence.

ACCESSION BF722070

VERSION BF722070.1 GI:12023063

KEYWORDS EST.

SOURCE house mouse.

ORGNISM Mus musculus

TAXID 10116

REFERENCE Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
NCBI-GAP <http://www.ncbi.nlm.nih.gov/ncbi/gap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strunberg, Ph.D.  
Email: rgs@b-fem11.nih.gov  
Ph. Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldino  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCBI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/ILNI at:  
[image.llnl.gov/image/html/tresources.shtml](http://image.llnl.gov/image/html/tresources.shtml)

NCI1170220

This is a reversed clone; similarity on wrong strand

Post primary sequencing stop: 356.

High quality sequence stop: 356.

FEATURES

CDS...  
Location/Qualifiers

1..634







ACCESSION AL044711  
 VERSION AL044711.1 GI:5432924  
 KEYWORDS EST  
 SOURCE human  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 491)  
 AUTHORS Pousetka A., Klein M., Mewes H.W., Gassenhuber J. and Wiemann S.  
 TITLE EST (Pousetka et al.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Pousetka A.J.  
 Department Lehnrich  
 Max-Planck-Institute for Molecular Genetics  
 Thomasstrasse 47, 10195 Berlin, Germany  
 Tel: +49-30-84131128  
 Fax: +49-30-84131128  
 Email: pousetka@mpg-berlin-dahlem.mpg.de  
 This is the 5' sequence of the clone insert  
 from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by DKFZ (German Cancer Research Center,  
 Heidelberg/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.  
 Note: This clone is a 5' truncated clone.  
 This clone (DKFZ03482235) is available at the RZPD in Berlin.  
 Please contact the RZPD: Resourcenzentrum Humankodex 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..491  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZ03482235"  
 /db\_xref="GeneID:10090"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pSport1; site\_1: NotI; site\_2: SalI"  
 BASE COUNT 117 a 127 c 147 g 100 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 45.00 Length: 45  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-750-590-2 x AL044711 ..  
 Align seg 1/1 to: AL044711 from: 1 to: 491  
 64 SerTysGlyAanLeuGluGlyAanAalaLeuGluGluGlyGlyAa 80  
 ||||||||||||||||||||||||||||||||||||||||||||  
 337 TCNAAGGCGAATCTGACTGTTGATCCCTTACCTTATACATGAGCTCA 406  
 80 pillethrhTSeTAspPThrAlaGlyTAaAaAalaLeuAlaAlaLc 97  
 ||||||||||||||||||||||||||||||||||||||||||||  
 407 TATTTCACACGAGCACTGACGAGGAGAAAGCTCTTACCTGAGCTCA 456  
 97 yATYtGlyhTAlaLeuGluGluGluGluGluGluGluGluGlu 108  
 ||||||||||||||||||||||||||||||||||||||||||||  
 457 ACTATGACATGCAATTTGGCTTACAAACTTCA 491  
 seq\_name: gb\_est2:BF148472 .  
 seq\_documentation\_block: 352 bp mRNA EST  
 LOCUS BF148472.1 NC1 CGAP\_Mam5 musculus cDNA clone IMAGE:3665994 3'  
 DEFINITION yBAB10.X1 NC1 CGAP\_Mam5 musculus cDNA clone IMAGE:3665994 3'  
 similar to TR:09Y375 09Y375 HYPOTHETICAL 59.8 kb PROTEIN ;, mRNA  
 sequence.  
 ACCESSION BF148472  
 VERSION BF148472.1 GI:11029867

KEYWORDS EST.  
 SOURCE mouse mouse  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 352)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: carbs@remail.nih.gov  
 Tissue Procurement: Lohrer Hemphausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 Image:1nl1.gov/image/html/resources.shtml  
 MGI:1426762  
 Possible reversed clone: polyT not found  
 Seq primer: -40UP from Glibco  
 High quality sequence stop: 331.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..352  
 /organism="Mus musculus"  
 /taxon="NC:780.10090"  
 /db\_xref="IMAGE:3665994"  
 /clone="lib" NCI-CGAP\_Mam5"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /note="Origin: mammary; Vector: pCMV-Sport6; site\_1: SalI;  
 site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lohrer Hemphausen/Rodin Humphreys,  
 NIH"  
 BASE COUNT 83 a 94 c 87 g 87 t 1 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 44.00 Length: 44  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-750-590-2 x BF148472/rev ..  
 Align seg 1/1 to reverse of: BF148472 from: 1 to: 352  
 1349 LeuGlnGlnGlnLeuAlaAaAaAaAaGlnGlnGlnGlnGlnAla 1365  
 ||||||||||||||||||||||||||||||||||||||||||||  
 280 CTGCAGCAGCAGCTGATGCTGATGACAGACAGCAAAAGTTATCC 231  
 1365 alleTyaTgThrHlsLeuLeuSerAlaAlaGlnGlnGlnAlaMeTAspAla 1382  
 ||||||||||||||||||||||||||||||||||||||||||||  
 230 TATCTACCGGACACCTCTGACGTCGCCACGACGACATGATGATG 181  
 1382 gpaTATGAlaAlaLeuLeuGlnGlnGlnGlnGlnGlnGlnGln 1392  
 ||||||||||||||||||||||||||||||||||||||||||||  
 180 ATCTGCAGCAGCAGCTTACATCCATCATCAG 149  
 seq\_name: gb\_est2:BG920973  
 seq\_documentation\_block: 605 bp mRNA EST  
 LOCUS BG920973 605 bp mRNA EST  
 DEFINITION 60822893F1 NC1 CGAP\_Mam6 Mus musculus cDNA clone IMAGE:4951775 5',  
 mRNA sequence.  
 ACCESSION BG920973



VERSION BG920973.1 GI:14301449  
 EST.  
 SOURCE house mouse.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 609)  
 NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS NIH-MGC  
 TITLE Unpublished (1999)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgsb@remail.nih.gov  
 Tissue Procurement: Gillbert Smith, Ph.D.  
 cDNA Library Prepared by: The I.M.A.G.E. Consortium, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHAM1008 row: 1 column: 24  
 High quality sequence stop: 608.  
 Location/Qualifiers  
 .. 609  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="4951775"  
 /clone\_lib="NCI\_CGAP\_Mam6"  
 /tissue\_type="mammary"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 156 a 163 c 112 t  
 ORIGIN

alignment\_scores:  
 Quality: 42.00 Length: 42  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x BG920973 ..

Align seg 1/1 to: BG920973 from: 1 to: 609

100 HsAlaLeuGcysLeuGlaLysLeuLeuGlnTyrAsnCysProThrGluH 116  
 332 CACGGCTGTCTCCCTGCAAACTCTACAGTACACACTCTCCACGCGCA 381  
 116 gvaLaAspLeuGlnGlyArGThAlaLeuHlaaPaPaLaHlaaMeAlaAspC 133  
 382 TGTGGAACTTCAGAGAAAGAACTGCGCTTCATGATGACCAAGCAACT 431

133 ysrProSerSerLeuGlnLeuLeuG 141  
 432 GTCCCTCCAGCAAGCACTGCTCTGT 457

seq\_name: gb\_est2:BI105199  
 seq\_documentation\_block:  
 LOCUS BI105199 925 bp mRNA  
 DEFINITION 6028934473f1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5038363 5',  
 mRNA sequence.  
 ACCESSION BI105199  
 VERSION BI105199.1 GI:14556092  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 925)  
 NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS NIH-MGC  
 TITLE Unpublished (1999)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgsb@remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Prepared by: The I.M.A.G.E. Consortium, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHAM1106 row: a column: 20  
 High quality sequence stop: 628.  
 Location/Qualifiers  
 .. 925  
 /organism="Mus musculus"  
 /strain="C57BL/6J (f1)"  
 /db\_xref="taxon:10090"  
 /clone\_image="5038363"  
 /clone\_lib="NCI\_CGAP\_Lu29"  
 /tissue\_type="spontaneous tumor; metastatic to mammary."  
 /stem\_cell\_origin="mammary"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 224 a 273 c 265 g 163 t  
 ORIGIN

alignment\_scores:  
 Quality: 42.00 Length: 42  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-750-590-2 x BI105199 ..

Align seg 1/1 to: BI105199 from: 1 to: 925

87 AlacGlytGAsnAlaLeuHsAlaLeuHsAlaLeuTyrGlyHsAlaLeuG 103  
 409 CCGCAAGAACTCTACAGTACACACTCTCCACGCGCAACTCTG 458  
 359 GCAGGAGGAGATCGCTTACCTGCGACGTACGTACGAGCAAGCGCTTGG 408

103 slsGlnLysLeuLeuGlnTyrAsnCysProThrGluHsAlaLeuG 120  
 120 InGlyArGThAlaLeuHlaaPaPaLaHlaaMeAlaAspC 128  
 459 AAGCAAGAACTGCGCTTCATGATGCA 484

seq\_name: gb\_est2:BI156388  
 seq\_documentation\_block:  
 LOCUS BI156388 546 bp mRNA  
 DEFINITION 61235866f1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5060099 5',  
 mRNA sequence.  
 ACCESSION BI156388  
 VERSION BI156388.1 GI:1616389  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 416)  
 NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS NIH-MGC  
 TITLE Unpublished (1999)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)



Thelating, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
CONTACT: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1800  
E-mail: msoares@wustl.edu  
This clone is available royalty-free through LBNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:504655

FEATURES  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 339.  
Location/Qualifiers  
1..348  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_1lb="Soares mouse BDMH"  
/sex="male"  
/tissue="type="heart"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="vector: p773B-Pac (pharmacia) with a modified  
polyA-polyA tail; Not I - oligo(dT) primer [5']  
3'1'; double-stranded cDNA was ligated to Eco RI adaptors  
(pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified p773 vector. RNA  
provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 101 a 96 c 98 g 53 t  
ORIGIN

alignment\_scores:  
Quality: 38.00 Length: 38  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-750-590-2 x AA499973 ..  
Align seg 1/1 to: AA499973 from: 1 to: 348

1289 ArgTleGlnGluSerAlaLysGlnIleGluAlaIysAspAsnIleTh 1305  
|||||  
58 AGGATCCAGGAGTCCGCCAGCAAGTCGAACGACAAATAGATTAAC 107  
1305 rGtUleUeuAsnAspValGluArgUeuLysGlnAlaIeuAsnGlyLeuS 1322  
108 TGAACCTGTGAAATGATCTCCAAAGACTTAACACGCCCTTCAAGCCCTTT 157  
1322 eGtInleuThrTyr 1326  
|||||  
158 CTCACCTCACCTCAC 171

seq\_name: qb\_esc2:BF147659  
seq\_documentation\_block: 603 bp mRNA EST 26-OCT-2000  
LOCUS BF147659  
DEFINITION IMAGE3169275.3 similar to TR:093735 O9Y3T5 HYPOTHETICAL 39.8 KD  
PROTEIN / mRNA sequence.  
ACCESSION BF147659  
VERSION BF147659.1 GI:11028981  
KEYWORDS EST.

SOURCE  
ORGANISM  
house mouse.  
Mus musculus.

REFERENCE  
AUTHORS  
TITLE  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
National Cancer Institute (1997)  
JOURNAL  
OTHER ESTS: us37f02.y1  
CONTACT: Robert Strausberg, Ph.D.  
Email: c94bps-r@mail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LBNL)  
DNA Sequencing by: Washington University School of Medicine  
Genetics Center  
Clone library distribution information can be  
found through the I.M.A.G.E. Consortium/LBNL at:  
[image.llnl.gov/image/hnml/lresources.shtml](http://image.llnl.gov/image/hnml/lresources.shtml)

FEATURES  
MGI:1064735  
Possible reversed clone: similarity on wrong strand  
Possible reversed clone: polyA not found  
High quality sequence stop: 494.  
Location/Qualifiers  
1..603  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_1lb="Soares, NMEBA, branchial\_arch"  
/tissue="type="branchial arches"  
/dev\_stage="embryo, 10.5 dpc"  
/lab\_host="DH10B (phage resistant)"  
/note="vector: p773B-Pac (pharmacia) with a modified  
polyA-polyA tail; Not I - oligo(dT) primer [5']  
3'1'; double-stranded cDNA was ligated to Eco RI adaptors  
(pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified p773 vector. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 178 a 161 c 170 g 93 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 38.00 Length: 38  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-750-590-2 x BF147659 ..  
Align seg 1/1 to: BF147659 from: 1 to: 603

1289 ArgTleGlnGluSerAlaLysGlnIleGluAlaIysAspAsnIleTh 1305  
|||||  
260 AGGATCCAGGAGTCCGCCAGCAAGTCGAACGACAAATAGATTAAC 309  
1305 rGtUleUeuAsnAspValGluArgUeuLysGlnAlaIeuAsnGlyLeuS 1322  
310 TGAACCTGTGAAATGATCTCCAAAGACTTAACACGCCCTTCAAGCCCTTT 359  
1322 eGtInleuThrTyr 1326  
|||||  
360 CTCACCTCACCTCAC 373

seq\_name: qb\_esc2:BF611191  
seq\_documentation\_block: 946 bp mRNA EST 30-OCT-2000  
LOCUS BF611191  
DEFINITION 601759451P1 NCI-CCAP\_L429 Mus musculus cDNA clone IMAGE:3988621 5',





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2002, 14:52:41 : Search time 3231.95 Seconds

23489-715 Million cell updates/sec

Title: US-09-750-590-1

Sequence: 1 cagctgtgagcgcgagat.....atcaaaaaaaaaaaaaaa 4730

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :  
1: gb\_db:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_com:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_da:\*  
14: gb\_fm:\*  
15: em\_da:\*  
16: em\_fm:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_com:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgo\_hum:\*  
31: em\_hgo\_in:\*  
32: em\_hgo\_rod:\*  
33: em\_hgo\_hum:\*  
34: em\_hgo\_in:\*  
35: em\_hgo\_rod:\*  
36: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score   | Query Match | Length | DB ID     | Description         |
|------------|---------|-------------|--------|-----------|---------------------|
| 1          | 3397.2  | 71.8        | 4428   | AF322916  | AF322916 Homo sapi  |
| 2          | 3387.8  | 71.6        | 4429   | AF322915  | AF322915 Bos tauru  |
| 3          | 3356.2  | 71.0        | 4643   | CFC3VS    | X93145 Canis fam1   |
| 4          | 2754.4  | 58.2        | 5520   | AB046781  | AB046781 Homo sapi  |
| C          | 1251.23 | 44.9        | 148041 | AC087699  | AC087699 Homo sapi  |
| G          | 125.5   | 28.0        | 155750 | AC010006  | AC010006 Homo sapi  |
| 9          | 125.2   | 28.0        | 155750 | AC010006  | AC010006 Homo sapi  |
| 8          | 125.2   | 28.0        | 155750 | AC010006  | AC010006 Homo sapi  |
| C          | 564.4   | 11.9        | 83775  | AC022466  | AC022466 Homo sapi  |
| 10         | 342.4   | 7.2         | 441    | AX193084  | AX193084 Sequence   |
| C          | 302.8   | 6.4         | 418    | G36547    | G36547 SHOC-56633   |
| 11         | 302.8   | 6.4         | 418    | G36547    | G36547 SHOC-56633   |
| 12         | 294.4   | 6.2         | 371    | AX071367  | AX071367 Sequence   |
| 13         | 259.8   | 4.5         | 401    | AC009269  | AC009269 Homo sapi  |
| 14         | 232.2   | 4.9         | 172850 | AC009269  | AC009269 Homo sapi  |
| 15         | 230     | 4.9         | 421    | AX192628  | AX192628 Sequence   |
| 16         | 230     | 4.9         | 421    | AX192628  | AX192628 Sequence   |
| 17         | 222     | 4.7         | 4925   | AF155135  | AF155135 Sequence   |
| 18         | 222     | 4.7         | 5043   | AB037755  | AB037755 Homo sapi  |
| 19         | 218.4   | 4.6         | 3024   | AF2020315 | AF2020315 Mus muscu |
| 20         | 218.4   | 4.6         | 4879   | AF2274866 | AF2274866 Mus muscu |
| C          | 159.8   | 3.4         | 250    | G14728    | G14728 human STR S  |
| 22         | 150     | 3.2         | 1673   | HS0800052 | AL050011 Homo sapi  |
| 23         | 147     | 3.1         | 2740   | AK023135  | AK023135 Homo sapi  |
| 24         | 137.2   | 2.9         | 2407   | AK014142  | AK014142 Sequence   |
| 25         | 137.2   | 2.9         | 2407   | AK014142  | AK014142 Sequence   |
| 26         | 117.6   | 2.5         | 231912 | AC087566  | AC087566 Homo sapi  |
| 27         | 103.4   | 2.2         | 42839  | AC027282  | AC027282 Homo sapi  |
| C          | 103.2   | 2.2         | 7218   | 166494    | 166494 Sequence 14  |
| 29         | 92      | 1.9         | 249282 | AF079430  | AF079430 Mus muscu  |
| 30         | 91.6    | 1.9         | 5420   | AF056936  | AF056936 Mus muscu  |
| 31         | 91.4    | 1.9         | 3148   | AF015461  | AF015461 Plasmodu   |
| 32         | 91.2    | 1.9         | 82234  | AC034104  | AC034104 Homo sapi  |
| C          | 89.4    | 1.9         | 182726 | AC008002  | AC008002 Drosophi   |
| 33         | 87.6    | 1.9         | 182726 | AC008002  | AC008002 Drosophi   |
| 34         | 87.6    | 1.9         | 182726 | AC008002  | AC008002 Drosophi   |
| 35         | 87.6    | 1.9         | 2065   | E10125    | E10125 DNA encodin  |
| 36         | 87.6    | 1.9         | 3399   | E10126    | E10126 DNA encodin  |
| 37         | 85.6    | 1.8         | 215287 | AC079432  | AC079432 Mus muscu  |
| 38         | 85.2    | 1.8         | 4995   | 3 PFAM58A | M69183 Plasmodu     |
| 39         | 85.2    | 1.8         | 6644   | E23356    | E23356 Virus vecto  |
| 40         | 85.2    | 1.8         | 7372   | E23357    | E23357 Virus vecto  |
| 41         | 85.2    | 1.8         | 7787   | E23355    | E23355 Virus vecto  |
| 42         | 85.2    | 1.8         | 7787   | E23355    | E23355 Virus vecto  |
| 43         | 85.2    | 1.8         | 7787   | E23355    | E23355 Virus vecto  |
| 44         | 84.8    | 1.8         | 11153  | AB002460  | AB002460 Homo sapi  |
| C          | 84.8    | 1.8         | 157566 | AC025534  | AC025534 Homo sapi  |

## ALIGNMENTS

RESULT 1  
AF322916 4428 bp mRNA PRI 15-FEB-2001  
DEFINITION Homo sapiens uveal autoantigen mRNA, complete cds.  
ACCESSION AF322916  
VERSION AF322916.1 GI:12240160  
KEYWORDS

SOURCE  
ORGANISM Homo sapiens  
Human.  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eularchia; Primates; Carnivora; Hominoidea; Homo.

REFERENCE  
Yamada K, Senju S, Nishimura T, Wurtz Y, Ishihara M, Nakamura S, Ohno S, Negi A and Nishimura Y.  
Identification of a novel autoantigen UACA in patients with

TITLE  
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1169-1176 (2001)  
PUBMED 11162650  
2 (bases 1 to 4428)

REFERENCE  
Yamada K., Senju S. and Nishimura Y.

SUMMARIES









|    |      |                                                                    |      |
|----|------|--------------------------------------------------------------------|------|
| QY | 971  | ggaicgagatcttggtctgaaagaatgcagtaagaatcttaatacaaaacggcgctgcgctg     | 100  |
| Dy | 668  | ggwtggcaatattgcttgacacgaatgcagtaagaatcttaattttaaattgcgtcgatata     | 727  |
| QY | 1031 | accttcgagcgccctctggcaatgaagaatctcttactatgcagaatctgtgcatactg        | 1090 |
| Dy | 728  | accttcgagcgccctctggcaatgaagaatctcttactatgcagaatctgtgcatactg        | 787  |
| QY | 1091 | gaactcttaactctctgaagatcgatctgaagaatctccaaagaatggagaaatcttg         | 1150 |
| Dy | 788  | gacattcttcaactcttgatgaagacgctacatgcgaataatgcgaataaagsgagaaatcttg   | 847  |
| QY | 1151 | aagaagaagccacactllacaaagaagaaatctctccagaatgcagatgaagaataacgc       | 1210 |
| Dy | 848  | aaagaagaagccacactllacaaagaagaaatctctccagaatgcagatgaagaataacgc      | 907  |
| QY | 1211 | aaacgaagatctggagagagaaatcaaacacttcagatctgagatctgaagaatgaagatcg     | 1270 |
| Dy | 908  | aaacgaagatctggagagagaaatcaaacacttcagatctgagatctgaagaatgaagatcg     | 967  |
| QY | 1271 | aaagaagaatctggagagaaatctccagaagaagaagaatctatctgtctaaagatctgcgc     | 1330 |
| Dy | 968  | aaagaagaatctggagagaaatctccagaagaagaagaatctatctgtctaaagatctgcgc     | 1027 |
| QY | 1331 | ctacagctaaagctgaagaggaagatgaatgctgtcgtatgcactctggaagatgcgaagaag    | 1390 |
| Dy | 1028 | ctacagctaaagctgaagaggaagatgaatgctgtcgtatgcactctggaagatgcgaagaag    | 1087 |
| QY | 1391 | aaagctatgccttggaagacgaagaatgatgctgcctcagatctgaagaatgcgaagaatc      | 1450 |
| Dy | 1088 | aaagctatgccttggaagacgaagaatgatgctgcctcagatctgaagaatgcgaagaatc      | 1147 |
| QY | 1451 | gaagcgctcgaagaagcagattcaatctctgaagatgcagatcttcagatccagaagatcat     | 1510 |
| Dy | 1148 | gaagcgctcgaagaagcagattcaatctctgaagatgcagatcttcagatccagaagatcat     | 1207 |
| QY | 1511 | ttccgcaagaatctatgcctctctcaacaagaatccaaatgataatgcagactca            | 1564 |
| Dy | 1208 | ttccgcaagaatctatgcctctctcaacaagaatccaaatgataatgcagactca            | 1267 |
| QY | 1565 | cagcgaaatctccacagcgcaatgcagatcttcagaaacggatctatgctgaagaacgag       | 1624 |
| Dy | 1268 | cagcgaaatctccacagcgcaatgcagatcttcagaaacggatctatgctgaagaacgag       | 1327 |
| QY | 1625 | gaagctagctcttaactcaatcaagctctatctcggaaagaagaatcttcaagaagaagatctg   | 1684 |
| Dy | 1328 | gaagctagctcttaactcaatcaagctctatctcggaaagaagaatcttcaagaagaagatctg   | 1387 |
| QY | 1685 | gagcaagaatgaagaactctctgattctgaagaagaagaacggcgctcaactccaaatgag      | 1744 |
| Dy | 1388 | gagcaagaatgaagaactctctgattctgaagaagaagaacggcgctcaactccaaatgag      | 1444 |
| QY | 1745 | ctggcgctcccaagaatctggagatgcagaagcctctgaagatctgaagaagtggaagaag      | 1804 |
| Dy | 1448 | ctggcgctcccaagaatctggagatgcagaagcctctgaagatctgaagaagtggaagaag      | 1507 |
| QY | 1805 | gattcttgatgaagaatcaaaagaacataagaatctgcctcttgaaagcgtctgcaagaagatg   | 1864 |
| Dy | 1508 | gattcttgatgaagaatcaaaagaacataagaatctgcctcttgaaagcgtctgcaagaagatg   | 1567 |
| QY | 1865 | taagaatcggagaagtaagaagaacaatgaaacaactcttcctgcctgcgaagaagacac       | 1924 |
| Dy | 1568 | taagaatcggagaagtaagaagaacaatgaaacaactcttcctgcctgcgaagaagacac       | 1627 |
| QY | 1925 | cttgcaatgaatgcagccttcaggaacacagcctgataagagaactgaagaatcgaatctga     | 1984 |
| Dy | 1628 | cttgcaatgaatgcagccttcaggaacacagcctgataagagaactgaagaatcgaatctga     | 1687 |
| QY | 1985 | aaagaacatgaagaatcaaatctagaagatctgcctgcgcaagaatgcggagaatctgaagaacaa | 2044 |
| Dy | 1688 | aaagaacatgaagaatcaaatctagaagatctgcctgcgcaagaatgcggagaatctgaagaacaa | 1747 |









```

OY 1285 aaabctcagcaagacagagaaatatacttgataaagctaaatgatttcagctaacct 1344
Db 2080 AAAATTGACAGAAACAAAGAAATCTTTGGATAAATCATATGGTTTCACTGATACACT 2139
OY 1345 gaatgagaagaaatgattgagctgataagcttgaaaatgaaagaaagctgaaagtcct 1404
Db 2140 GAATGAGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2199
OY 1405 ttctgagcccaagaaagaaagcaatgaaagaaagcaatgaaagcaatgaaagcaatgaaag 1464
Db 2200 TTGGCGAGCTAAAGAAACGACATGAAGAAAGCTTAAAGACATATGACGCTTCAAAA 2259
OY 1465 tgaattaaagattttgagagtgatacttgaaatgaaagaaagcaatgaaagcaatgaaag 1518
Db 2260 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2319
OY 1519 aagaaatgagctgcttgaaagaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 1578
Db 2320 AAGAAATGAGCTCTTTAAACAGCGATGATATATGAGACGACATCAATGATCTTCCG 2379
OY 1579 aagcaatgagctgataatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 1638
Db 2380 AGCTATACAGCCCAATGAGAAACGACATGATGATGATGATGATGATGATGATGATGATGAT 2439
OY 1639 taatcagagctgcttgaaagaaagcaatgaaagaaagcaatgaaagcaatgaaagcaatgaaag 1698
Db 2440 CAGTCAAAGCTCATCTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2499
OY 1699 ttctctgatactgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 1758
Db 2500 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2559
OY 1759 gtagagagtgagagagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 1818
Db 2560 GCGAGATGCAAGCTTTACGATTAAGATGAGAAAGCTTAAAGAGATGAGATGAGATGAGATGAG 2619
OY 1819 gatataagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 1878
Db 2620 GATATAAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2679
OY 1879 taatgagaaagaaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 1938
Db 2680 TAAAGTTAAACAAATGCGACGCCATTTCTTCCGCTTAAAGAAAGCTTAAAGAGATGAGATGAGAT 2739
OY 1939 gtagagagtgagagagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 1998
Db 2740 AGCTATACAGCCCAATGAGAAACGACATGATGATGATGATGATGATGATGATGATGATGAT 2799
OY 1999 gaaatcagagtgagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 2058
Db 2800 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2859
OY 2059 aatgagtgagagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagag 2118
Db 2860 GATATAAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2919
OY 2119 gtagagagtgagagagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 2178
Db 2920 ACGAAGAGACTTACTATCTGTAAATGAGAAAGCTTAAAGAGATGAGATGAGATGAGATGAGATGAG 2979
OY 2179 gaatgagaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 2238
Db 2980 GATGAGAGCCCAATGAGAAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3039
OY 2239 atttgaagaaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaat 2298
Db 3040 ATTTGAAACATGAAAGCTCATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3099
OY 2299 agatgagaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaag 2358
Db 3100 AGAAATGAGAAAGCAATGAGAAACGACATGATGATGATGATGATGATGATGATGATGATGATGAT 3159
OY 2359 tgaagattgaaagcaatgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 2418

```

---

```

Db 3160 TGAAGATGTAAGCCAAAGCTTGTCTGACGATGAAACGAGCAAGCAAGCAAGCAAGCAAGCAAG 3219
OY 2419 gtagagagtgagagagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 2478
Db 3220 GAGCGATTAAACAGAAATGAGAGCAAGCTTGGAAAGAAATGATGATGATGATGATGATGATGAT 3279
OY 2479 aaatcagagtgagagagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaat 2538
Db 3280 AAATTAAGCAATTAAGAAATGAGAAACGACATGATGATGATGATGATGATGATGATGATGATGAT 3339
OY 2539 aaatcagagtgagagagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaatgaaagcaat 2592
Db 3340 GCAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT 3399
OY 2593 aagaaatgaaagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 2652
Db 3400 AGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3459
OY 2653 gtagagagtgagagagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 3518
Db 3460 AGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3579
OY 2713 ttctgaatgaaagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 2772
Db 3520 CTTAAGTAAAGCAAGCCGCTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3579
OY 2773 aagaaatgagctgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagag 2832
Db 3580 AGCAATGAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3599
OY 2833 taataaagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 2892
Db 3640 GAAATTAAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3699
OY 2853 gtagagagtgagagagctgagctgagagtgagagtgagagtgagagtgagagtgagagtgagag 2952
Db 3700 GAGAAATGAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3759
OY 2953 ctgagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagag 3012
Db 3760 ACTGAATGAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3819
OY 3013 gtagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagag 3072
Db 3820 AGATTAATTAAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3879
OY 3073 gtagagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 3132
Db 3880 GGAATTAAGCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3939
OY 3133 gatgagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagag 3192
Db 3940 GATGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 3999
OY 3193 taatgagaaagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 3252
Db 4000 CAGCTATGAGAAAGCAATGAGAAACGACATGATGATGATGATGATGATGATGATGATGATGATGAT 4059
OY 3253 gtagagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 3312
Db 4060 GAGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 3372
OY 3313 gtagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagag 3372
Db 4120 TGAGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 4179
OY 3373 gtagagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 3432
Db 4180 GAGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 4239
OY 3433 caagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 3492

```







```

Db 126784 TTTAAAGGAAAGAACTGGGAAAAACCTCAACAACTTAAAGCTGAGTCAATCAAGCTGGCC 126725
Oy 3119 ggcacatgaagaagaatgagtggtccaaagagatgaaagagccaaagcaacagc 3178
Db 126774 GAGCAGACAGGAAAGATACACTGCTCAATCAAGATGAGAAAGGTGACAGATCAATAT 126665
Oy 3179 gctcgaatactcgtcctgaatgacaaaagaagagagaggtgtccctcctgagag 3238
Db 126664 GCTGAATCTTGCCCACTACAGAAAAGCCAAAGAGATGTGTCACTCAATCCAGCA 126605
Oy 3239 atctgagcccgaaagagagagactcgacagctcgcagagatcacaagctaaatctc 3298
Db 126604 ATTAAACCCCAAGAAAGAGAGCTGACAAATCAAGAAATCAATTAAGGTAAATACCC 126545
Oy 3299 cggatctacgcctctggagaggtgtcgagagaaatctaaagccactgagaaagactaaa 3358
Db 126544 CCATTTGCTACCTTTGAGAGCTGCGAAGAAATTTGAGAGAGAGAGAGAGAGAGAG 126485
Oy 3359 gaacacatctcccgacagacagaaagataataacagatgaagagagagagagagag 3418
Db 126484 GAGCAGTATACAGACAGACACAAAGTGTGTGCTAGTAAAGAAAGATCAAGAAATAC 126425
Oy 3419 aagcaagaagaatgacaaagttaaagaagaagatccctcctcctcgcagagagatc 3478
Db 126424 AAGCAAGAGAGATGACAGATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126365
Oy 3479 aagagatctacatgagaaatctctctgaaagagagagagcttaagcgaaagaaag 3538
Db 126364 AAGCAGAGAGATGACAGATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126305
Oy 3539 gaggctgaacagacagctctaaagaagcctgtgtgcaagaatcacacagagcgcaag 3598
Db 126304 GAGCTCAACAGAGAGATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126245
Oy 3599 gaggagcctcgtgagagaaatgccaagcagagcctctgaagctcctcgaagcagac 3658
Db 126244 GAGAGAGCTGTGAAGAGAAATCCCAAGAGAGCTTGTGAGATCTGAGATCTGAG 126185
Oy 3659 ttgcgaagcagcagcttcctcgtgagagagcttgagctccctgaaagaaatcct 3718
Db 126184 TTGCAAAAACAGATGTCATTTGAGAAAGGTGAGAGCTCTGATAAAATCTTTATG 126125
Oy 3719 acatctgagacactcaagagagacataaactcaagacagagatctcagagaaagagc 3778
Db 126124 ACATTTGAAATCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126065
Oy 3779 cagcagcgagacccacgtcgagacatgctcgtgagaaatcagaaagacccctcgt 3838
Db 126064 CAGCAG 126005
Oy 3839 gctcgaagcttgcagaggtctaaagaagcctctgagaaagagttgagatcaataaag 3898
Db 126004 GCAGACATCTTGCAGATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 12845
Oy 3899 ttgagagaaagaagaagaagaagacaaacaaactgaagaggtcccaacacagact 3958
Db 125844 TTGAG 125885
Oy 3959 gaggctcgaagatctcaaacagcgtctaaagaaatctgagacgtggaggtgtgact 4018
Db 125884 GAGCTTCAGAAATCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128635
Oy 4019 tgcgaatctaaagcaagaaagcgtcttgagagacaaatctcgactcaagagaaag 4078
Db 125824 TCTAATATTAACAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123765
Oy 4079 ttggcgaatctgagatgagaaagatgagaaagatgagagagaggttttcgtcccaaa 4138
Db 125764 TTGGCCATCTCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123705
Oy 4139 aaggaacgtcgtcgtcaaaagtgaagagatgctcctcactctgctgctgagcagaaag 4198

```

```

Db 125704 AAGCAATATCTGCAAAAAGATGAGAGAGATTTACGCTATTCACATTCAGCAGAGAAAT 125645
Oy 4199 aagatctcagagagacgtgagacaaatctcttaacaaacatcgaagacatcagagaa 4258
Db 125644 AAGCATCAG 125585
Oy 4259 atacagaaatctgcacaaacaaatctgagaaagagatcaagata 4303
Db 125584 ATTCAGAAATCTCTTAAACAAATTAAGAGAGAGAGAGAGAGATTAAGATA 125440

RESULT 6
AC010076/c DNA HNG 01-SEP-2000
LOCUS Homo sapiens chromosome 15 clone RP11-64X10, WORKING DRAFT
DEFINITION SEQUENCE, 19 ordered pieces.
ACCESSION AC010076
VERSION AC010076.5 GI:9957994
AUTHORS HUGO, HUGO, PHASE2: HUGO, DRAFT.
SOURCE Human
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 155760)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A.,
Resnick, R., Treicoff, R. and Hood, L.
Sequencing of human chromosome 15 D15S114-D15S115 region
Submitted (11-SEP-1999) Multimegabase Sequencing Center, University
of Washington, PO Box 357730, Seattle, WA 98195, USA
On Sep 11, 2000 this sequence version replaced g1:8312412.
Accession number: AC010076
Direct Submission
Rowen, L., Madan, A., Qin, S., Abbadi, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S.,
Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.
Submitted (11-SEP-1999) Multimegabase Sequencing Center
Center: Multimegabase Sequencing Center
Center code: UMMS
Web site: http://chroma.mbl.washington.edu/msg/www
Contact: leeroensystemsbiology.org

----- Summary Statistics -----
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Insert size: 1500bp; average insert size 0.950399
Insert size: 15000bp; average insert size 0.950399
Quality coverage: 4.6x in Q20 bases; sum-of-coverage

----- Sequence Quality Assessment -----
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
one error per 1000 bases. Quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

----- NOTE -----
* NOTE: This is a 'working draft' sequence. It currently
consists of 19 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
of the gaps between them is given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
* 39366: contig of 39368 bp in length
* 39469: gap of unknown length
* 43256: contig of 4488 bp in length
* 43957: gap of unknown length

```

| Query Match           | 38.0%                                                             | Score 1796.2                                                      | DB 2           | Length 155760 |
|-----------------------|-------------------------------------------------------------------|-------------------------------------------------------------------|----------------|---------------|
| Best Local Similarity | 77.5%                                                             | Pred. No. 0                                                       |                |               |
| Matches 2311          | Conservative                                                      | 0                                                                 | Mismatches 428 | Indels 242    |
|                       |                                                                   |                                                                   |                | Gaps 4        |
| 1565                  | cagctgaacttccacagcagcagcagccgactcatttcgaacagccagatctctgttaagacggt | 1524                                                              |                |               |
| Db 135038             | CAGCTGACTTCTCCCAAGATACCAAGCCCATTCCTGCAACGATCTAGTAAAGCCTCTG        | 134979                                                            |                |               |
| Qy                    | 1625                                                              | gagcctgacctaccacaaacaaacccctcatttcgaaacaaagaattttaaagaagatla      | 1684           |               |
| Db 134978             | GAGCCTGACCTTACCCAGCAACGCAACGTCATCTCTGAAATATGAAATTTTAAAGAAAGTTTA   | 134919                                                            |                |               |
| Qy                    | 1685                                                              | gagagcaagaggaactctcgtcgatctgacgaacaaacaaagacagacccaactccaatgaa    | 1744           |               |
| Db 134918             | GAGAGCAATGCCAACTCTTCTGTGTCGCAACAAAGACGACGACGTAAAGCTCCAAATTTAA     | 134659                                                            |                |               |
| Qy                    | 1745                                                              | ctcgagctcaaaagcttgcgagagtggaacagcctctgacatctggaatctggaagag        | 1804           |               |
| Db 134858             | CTCGCAACCAAAAGTCGCAGATGCAAAAGCTTTACACTTAGAATGTGAAAGGCTTAAGGAG     | 134799                                                            |                |               |
| Qy                    | 1878                                                              | gattcagatgagcagaataaagaacacatcagagaatgactctggaagacagcgtgcgaagaaab | 1864           |               |
| Db 134798             | GATTCAATGAGACACATTTGATGATGACAGATCTCTTAAGAAATGTGCGCAAGAGAGG        | 134739                                                            |                |               |
| Qy                    | 1885                                                              | tattgattcgagagataaagtgaaacaaagagagacactctctgctctgaaagagac         | 1924           |               |
| Db 134738             | TATTGATCGACAGAGCTAAAGTTAAACAAATGCGAGCCATTTTCTTGCCCTTAAAGAAC       | 134679                                                            |                |               |
| Qy                    | 1925                                                              | ctcgacagagagctgagccaccttggaacaccagagctgacggaagaaactcgaagatcagttg  | 1984           |               |

[illegible]



```

/def-stage="embryo, 10 weeks"
/issue-type="whole embryo, mainly head"
/notes="cloning vector: pME18SFJ3"
22. .s1628
/notes="unnamed protein product"
/record_status="new"
/submitter="JGI-TMA91457.1"
/db_xref="gi:7021993"
/translation="MRRREKSELSERINQKRELVNKKALQHQPEHHVOVSRLEQ
KSGGELGKLTFLNKTQKTEIEVYLDLNKKLLEQDAINTEKNNIYVPLKSGEDMK
KSHALITIDILNRLKLIDYQKTEKRLKEMKLLLEDSLSQDSRLTEVYVPPKHEK
ITAKLSVITLQKLOSTELKSGCEQDQK IKAELISPTNLLKKMSQVNOVYVTEH
TLNQTILANKEILQVLSKKKREDFI NOEYV IHNQTEKLRNLTNOV INAEY ISLAEY
ENKMSLSQSSKRVQDSKAEITLANTIRQCEETIYLAIVIAKQKELDTIOGCIKRYV
EYKNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQV
PKNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQV
VONLQHQPEHHVOVSRLEQKSGGELGKLTFLNKTQKTEIEVYLDLNKKLLEQDAIN
TEKNNIYVPLKSGEDMKKSHALITIDILNRLKLIDYQKTEKRLKEMKLLLEDSLSQ
DSRLTEVYVPPKHEKITAKLSVITLQKLOSTELKSGCEQDQK IKAELISPTNLLKK
MSQVNOVYVTEHITLNQTILANKEILQVLSKKKREDFI NOEYV IHNQTEKLRNLTNO
V INAEY ISLAEY ENKMSLSQSSKRVQDSKAEITLANTIRQCEETIYLAIVIAKQKEL
DTIOGCIKRYV EYKNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSK
QVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQVNSKQV
NSVPLQHQPEHHVOVSRLEQKSGGELGKLTFLNKTQKTEIEVYLDLNKKLLEQDAIN
SEK"
707 a 236 c 348 g 337 t

```

```

Query Match 100.0%: Score 1215.6: DB 9: Length 1628:
Best Local Similarity 84.6%: Pred. No. 1:5e-238:
Matches 1378: Conservative 0: Mismatches 244: Indels 6: Gaps 1:

QY 2282 aaggcaaaaataataatagaatgtagaaagaaatctgaanaagtcacattaaagaactaga 2341
Db 1 AAGCCAAAAAATTTAGTNGAATCGAAGAACAAACATCGAAAAATCTATGTGGAATTGCA 60
QY 2342 cactaaagagagaaacttgaagatttgaaggccaacttgctcagaagtcacaaacagag 2401
Db 61 CAGTAAAGACAGACATCTGAGACATTTAGAGCCAGACAGCTGCTACACACAGAGAC 120
QY 2402 gaataatgagacactgaagaagcagaatttgaagcaagaatctggaaactctggaaagc 2461
Db 121 GAAATGACACAGGTTAAGACAGATTTGAAACAAATATACGAACTTGCGAAGAACATC 180
QY 2462 acctgagcttaacacagaaataataagctctcaaaaagaaatcgaaaaggtctctcgagat 2521
Db 181 ACTGATTTAACTATTGAAAAATACACACACACAAAAAGAAATTTGAAAATTTTGTGAT 240.
QY 2522 aatgaactctctccacacagataaacttaacttaacaaatgaaagcaagaaa-----atgctc 2575
Db 241 AATGACCTCTCCACAGCAAGCAACCATTAATTAACAAATTAATGAAGAAATTAATTTAT 300
QY 2576 cctttaaagctaaagtgaagaaatgaaaagtcacatgagtcgaatctgtatgatttgact 2635
Db 301 CCGTTAAAGTAACTGAACACATCGAAGAGTCACATCTGCAATTTATATGACATCTTAA 360
QY 2636 aaaaagactctgaagatgagacacaaatatcagaagaaagaatttgaagattggagaagatg 2695
Db 361 ACAAACCTTTTAAATGCTAACACAAAATTTATCGAAAAGAAATTTGCAATTCGAGAAAT 420
QY 2696 ctatctggaaatgctcagattttaaagtaaaatgctcagcccttggaagctgctctcatatc 2755
Db 421 CTAATCGAAAATACACGCTTAATGTAAGATGTTAACGCCCTCGAAGCTGTTTGTGACCT 480
QY 2756 ccagagagaacacgaaaaaagaatgtagctctgaaatccaaatacaactgaacttaagaag 2815
Db 481 CCGTGAACACATGAAGAAAGAAATAGCTCTGAATCAATATGCTATGAACTTAAGAA 540
QY 2816 cagctctctgaacttaataaagaatgtagtgaagacaaagaaataataatcaactgaat 2875
Db 541 CAGCTCTCTGAACCTTAAAGAAAAATTTGTGTAAACACGAGAAATTAACGCTCTCA 600
QY 2876 tctgaaacaaatgattctgaaanaaaacaaatggagtcactcagatgtagtcgcgaatgaagccat 2935
Db 601 TGTGAAAACACTAACTTGAAGAAGATGATGAACTAGTATGATGCGCAATTTAAACCCAT 660
QY 2936 gaagagatttaaacctgactctggatgagcaatcttggaataaacaatagagatttaagatg 2995
Db 661 GAAAGAGTTAAATATGACATCGAATGACAGCTTACGCAAAATTAACAGAAATTAATTTAGAT 720

```

|            |            |                                                                |                                                                 |
|------------|------------|----------------------------------------------------------------|-----------------------------------------------------------------|
| Oy         | 2956       | gtaagaagaaagcgtgaagatctaaacacgaattcttgtaaaataaaagaatgagagcaaa  | 3035                                                            |
| Oy         | 721        | gtgaaagaaaaatttgaaagatgttaatacgaattttgttaaaataaaagatgaacatga   | 780                                                             |
| Oy         | 3056       | attctaaagaagatactggaagaaacctcaagacagctaaagcctggagatgacatcaga   | 3115                                                            |
| Oy         | 3116       | aaagaagatgagaagaagaagatgagacctaaagaagagcagaagaagagctcagaagac   | 3175                                                            |
| Oy         | 841        | gcagacgcacgcgcgaagagatgagcctccctgaatgcagagctgaaataaagctgcagat  | 900                                                             |
| Oy         | 3176       | agagctgaaatactgagctgaagtacaaataaagaacgcagagagatgctgacccctgacag | 3235                                                            |
| Oy         | 901        | aatgcctaaatctctgcgcacactcaabaaabagccaaabaaatcttgacatgcagctcc   | 960                                                             |
| Oy         | 3236       | ggagctctgagccgcgaagaagaagaacatccagacgaataagaagatgcatcagaatctaa | 3295                                                            |
| Oy         | 961        | gaattttaaagccccaagaaagagacatccacacaaacaaagattccattgaagtaattat  | 1020                                                            |
| Oy         | 3296       | gctccagatcatccagcttggaagaagcttggaagaagaatcttaagaccatctgaagaaga | 3355                                                            |
| Oy         | 1021       | gcccaattatgcctgagctgagacagctgacgaagacgaagaaattttaaagcagcaaa    | 1080                                                            |
| Oy         | 3356       | aaagaagatgacatccgcagctgaagacagctgataatactccatgagaagaagcagaag   | 3415                                                            |
| Oy         | 1081       | aaagacacattttatcgacgcagcgcacaaagaaatgactgcctgagaaagaaagacaaa   | 1140                                                            |
| Oy         | 3416       | tgcagaagaagaatgacgaagttaagaagaagatcctcaactctcagaagaatctaaag    | 3475                                                            |
| Oy         | 1141       | aaacacagcagaaagatgacacagctaaagaaagagacattttaccctctacaaagatttga | 1200                                                            |
| Oy         | 3476       | gttaaaatgcttccatcttgpaatctcttaagaagaagaagaagcttaagaagaaga      | 3535                                                            |
| Oy         | 1201       | gattaaacacgttccatttgagaaactcctcaatgaatgcgaagacatttgaacgaaana   | 1260                                                            |
| Oy         | 3536       | gaagaagcctgaaacagcagcttaaaagacctctgacgaagaatacagaagaagaagaag   | 3595                                                            |
| Oy         | 1261       | gcagacatgcataacacagcaattaaagacacgttgcacgaataacagaaagaaatgctg   | 1320                                                            |
| Oy         | 3596       | aaagaagaatgccttgagaagaatgacgaagctgagctgaatctgcagagcagaact      | 3655                                                            |
| Oy         | 1321       | aaagaaacactagtagaagaaatgcccaaacacacacttctgaaatgactgcagcaaaat   | 1380                                                            |
| Oy         | 3656       | cctctgcagaagaagacatgctccgcctgagaagaagctgagctccctcgaaaaaaactctg | 3715                                                            |
| Oy         | 1381       | ctttttgcaaaaaacacatgctcatttggaadagctttgacgtctgcaaaaaatctctta   | 1440                                                            |
| Oy         | 3716       | gtctcaacacagagacactcaagaagaagacatgaaacactgaagagatctctgaagaaga  | 3775                                                            |
| Oy         | 3776       | cagacagaagctgacacaaactctggcgacagctgacgtgacgtgacgaagaactcctctg  | 3835                                                            |
| Oy         | 1501       | cagcagacacgtgacacaaacgtcagatcatttgtagtgagaaacaaagaaactctctgt   | 1560                                                            |
| Oy         | 3836       | ctgagctgagacatctgacagtttaagaagaatcttggaagaagaagtctgaaatcataa   | 3895                                                            |
| Oy         | 1561       | ctgcacagacattttgcacatttgaagaagacatttgaagaaacaaagcttgaaatcatt   | 1620                                                            |
| Oy         | 3896       | agactcag 3903                                                  |                                                                 |
| Oy         | 1621       | acgttgag 1628                                                  |                                                                 |
| RESULT     | 8          |                                                                |                                                                 |
| AC022466   | AC022466   | 83775 bp                                                       | DNA                                                             |
| LOCUS      |            |                                                                | HTG                                                             |
| DEFINITION |            |                                                                | 04-FEB-2000                                                     |
|            |            |                                                                | Homo sapiens chromosome 15 clone CTD-302T24 map 15q23, LOW-PASS |
| SEQUENCE   |            |                                                                | SAMPLING                                                        |
| AC022466.1 | GI:6893623 |                                                                |                                                                 |
| VERSION    |            |                                                                |                                                                 |

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

HTG: HTGS\_PHA590.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 83775)  
Bowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,  
Bischoff, R., Dickhoff, R., Fleecewood, P., Harrison, G.,  
James, R., Kaur, A., Madan, A., Owen, M., Ratcliffe, A., Shaffer, T.  
and Hood, L.  
Sequencing of human chromosome 15 DIS146-D15S117 region  
Unpublished  
2 (bases 1 to 83775)  
Bowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,  
Bischoff, R., Dickhoff, R., Fleecewood, P., Harrison, G.,  
James, R., Kaur, A., Madan, A., Owen, M., Ratcliffe, A., Shaffer, T.  
Direct Submission  
Submitted (04-P88-2000) Multiphase Sequencing Center, University  
of Washington, PO BOX 357730, Seattle, WA 98195, USA  
-----  
Center: Multiphase Sequencing Center  
Center code: UWMSC  
Web site: <http://chroma.mbl.washington.edu/msg-www>  
Contact: Jeevaretnu, Washington.edu  
Sequencing strategy: Sanger, 1508752  
Chemistry: Dye-terminator Big Dye; 90% of reads  
Chemistry: Dye-terminator Big Dye; 10% of reads  
Assembly program: Phrap; version 0.990399

-----  
\* NOTE: This record contains 96 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the reads are not necessarily in order.  
\* Arbitrary low-pass sequences appear in contigs for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
887: contig of 867 bp in length  
888 1780: gap of unknown length  
1781: contig of 865 bp in length  
2645: contig of 885 bp in length  
2646 gap of unknown length  
3525: contig of 880 bp in length  
3526 gap of unknown length  
4401: contig of 876 bp in length  
5290: gap of unknown length  
5291: contig of 889 bp in length  
6155: gap of unknown length  
6156: contig of 865 bp in length  
7151: gap of unknown length  
7152: contig of 996 bp in length  
8036: gap of unknown length  
8037: contig of 885 bp in length  
8948: gap of unknown length  
8949: contig of 912 bp in length  
9836: gap of unknown length  
9837: contig of 888 bp in length  
10705: gap of unknown length  
10706: contig of 869 bp in length  
11579: gap of unknown length  
11580: contig of 874 bp in length  
12459: gap of unknown length  
12460: contig of 880 bp in length  
13321: gap of unknown length  
13322: contig of 862 bp in length  
14390: gap of unknown length  
14391: contig of 869 bp in length

14391 gap of unknown length  
15058: contig of 868 bp in length  
15059 gap of unknown length  
15923: contig of 865 bp in length  
15924 gap of unknown length  
16802: contig of 879 bp in length  
16803 gap of unknown length  
17680: contig of 878 bp in length  
17681 gap of unknown length  
18548: contig of 868 bp in length  
18549 gap of unknown length  
19413: contig of 865 bp in length  
19414 gap of unknown length  
20289: contig of 876 bp in length  
20290 gap of unknown length  
21156: contig of 867 bp in length  
22031: gap of unknown length  
22032: contig of 875 bp in length  
22916: gap of unknown length  
22917: contig of 885 bp in length  
22918 gap of unknown length  
23786: contig of 870 bp in length  
23787 gap of unknown length  
24653: contig of 869 bp in length  
25519: gap of unknown length  
25520: contig of 864 bp in length  
26383: gap of unknown length  
26384: contig of 864 bp in length  
27253: gap of unknown length  
27254: contig of 870 bp in length  
28119: gap of unknown length  
28120: contig of 866 bp in length  
28987: gap of unknown length  
29861: contig of 868 bp in length  
30737: gap of unknown length  
30738: contig of 876 bp in length  
30739 gap of unknown length  
31601: contig of 864 bp in length  
31602 gap of unknown length  
32481: contig of 880 bp in length  
32482 gap of unknown length  
33347: contig of 866 bp in length  
33348 gap of unknown length  
34218: contig of 877 bp in length  
34219 gap of unknown length  
35081: contig of 863 bp in length  
35082 gap of unknown length  
35948: contig of 867 bp in length  
35949 gap of unknown length  
36814: contig of 866 bp in length  
36815 gap of unknown length  
37687: contig of 873 bp in length  
37688 gap of unknown length  
38583: contig of 866 bp in length  
38584 gap of unknown length  
39452: contig of 869 bp in length  
39453 gap of unknown length  
39454 gap of unknown length  
40335: contig of 882 bp in length  
41201: gap of unknown length  
41202: contig of 867 bp in length  
42067: gap of unknown length  
42068: contig of 866 bp in length  
42933: gap of unknown length  
42934: contig of 866 bp in length  
43801: gap of unknown length  
44675: contig of 874 bp in length  
44676 gap of unknown length  
45538: contig of 863 bp in length  
45539 gap of unknown length

```

      45339      46405: contig of 867 bp in length
      46406      47263: contig of 864 bp in length
      47270      48139: contig of 870 bp in length
      48140      48959: gap of unknown length
      49000      49866: contig of 867 bp in length
      49867      50729: gap of unknown length
      50730      51588: contig of 859 bp in length
      51589      52460: gap of unknown length
      52461      53323: gap of unknown length
      53324      54189: contig of 866 bp in length
      54190      55053: gap of unknown length
      55054      55910: contig of 857 bp in length
      55911      56776: gap of unknown length
      56777      57669: contig of 866 bp in length
      57670      58539: gap of unknown length
      58530      59404: gap of unknown length
      59405      60263: contig of 859 bp in length
      60264      61140: gap of unknown length
      61141      61997: contig of 857 bp in length
      61998      62885: contig of 888 bp in length
      62886      63756: gap of unknown length
      63757      64624: gap of unknown length
      64625      65487: contig of 863 bp in length
      65488      66346: contig of 859 bp in length
      66347      67207: contig of 861 bp in length
      67208      68075: gap of unknown length
      68076      68943: gap of unknown length
      68944      69811: gap of unknown length
      69812      70679: gap of unknown length
      70680      71547: gap of unknown length
      71548      72415: gap of unknown length
      72416      73283: gap of unknown length
      73284      74151: gap of unknown length
      74152      75019: gap of unknown length
      75020      75888: gap of unknown length
      75889      76756: gap of unknown length
      76757      77624: gap of unknown length
      77625      78493: gap of unknown length
      78494      79361: gap of unknown length
      79362      80229: gap of unknown length
      80230      81097: gap of unknown length
      81098      81965: gap of unknown length
      81966      82833: gap of unknown length
      82834      83701: gap of unknown length
      83702      84569: gap of unknown length
      84570      85437: gap of unknown length
      85438      86305: gap of unknown length
      86306      87173: gap of unknown length
      87174      88041: gap of unknown length
      88042      88909: gap of unknown length
      88910      89777: gap of unknown length
      89778      90645: gap of unknown length
      90646      91513: gap of unknown length
      91514      92381: gap of unknown length
      92382      93249: gap of unknown length
      93250      94117: gap of unknown length
      94118      94985: gap of unknown length
      94986      95853: gap of unknown length
      95854      96721: gap of unknown length
      96722      97589: gap of unknown length
      97590      98457: gap of unknown length
      98458      99325: gap of unknown length
      99326      100000: gap of unknown length

```

```

      2391      2450: contig of 867 bp in length
      2451      2510: gap of unknown length
      2511      2570: gap of unknown length
      2571      2630: gap of unknown length
      2631      2690: gap of unknown length
      2691      2750: gap of unknown length
      2751      2810: gap of unknown length
      2811      2870: gap of unknown length
      2871      2930: gap of unknown length
      2931      2990: gap of unknown length
      2991      3050: gap of unknown length
      3051      3110: gap of unknown length
      3111      3170: gap of unknown length
      3171      3230: gap of unknown length
      3231      3290: gap of unknown length
      3291      3350: gap of unknown length
      3351      3410: gap of unknown length
      3411      3470: gap of unknown length
      3471      3530: gap of unknown length
      3531      3590: gap of unknown length
      3591      3650: gap of unknown length
      3651      3710: gap of unknown length
      3711      3770: gap of unknown length
      3771      3830: gap of unknown length
      3831      3890: gap of unknown length
      3891      3950: gap of unknown length
      3951      4010: gap of unknown length
      4011      4070: gap of unknown length
      4071      4130: gap of unknown length
      4131      4190: gap of unknown length
      4191      4250: gap of unknown length
      4251      4310: gap of unknown length
      4311      4370: gap of unknown length
      4371      4430: gap of unknown length
      4431      4490: gap of unknown length
      4491      4550: gap of unknown length
      4551      4610: gap of unknown length
      4611      4670: gap of unknown length
      4671      4730: gap of unknown length
      4731      4790: gap of unknown length
      4791      4850: gap of unknown length
      4851      4910: gap of unknown length
      4911      4970: gap of unknown length
      4971      5030: gap of unknown length
      5031      5090: gap of unknown length
      5091      5150: gap of unknown length
      5151      5210: gap of unknown length
      5211      5270: gap of unknown length
      5271      5330: gap of unknown length
      5331      5390: gap of unknown length
      5391      5450: gap of unknown length
      5451      5510: gap of unknown length
      5511      5570: gap of unknown length
      5571      5630: gap of unknown length
      5631      5690: gap of unknown length
      5691      5750: gap of unknown length
      5751      5810: gap of unknown length
      5811      5870: gap of unknown length
      5871      5930: gap of unknown length
      5931      5990: gap of unknown length
      5991      6050: gap of unknown length
      6051      6110: gap of unknown length
      6111      6170: gap of unknown length
      6171      6230: gap of unknown length
      6231      6290: gap of unknown length
      6291      6350: gap of unknown length
      6351      6410: gap of unknown length
      6411      6470: gap of unknown length
      6471      6530: gap of unknown length
      6531      6590: gap of unknown length
      6591      6650: gap of unknown length
      6651      6710: gap of unknown length
      6711      6770: gap of unknown length
      6771      6830: gap of unknown length
      6831      6890: gap of unknown length
      6891      6950: gap of unknown length
      6951      7010: gap of unknown length
      7011      7070: gap of unknown length
      7071      7130: gap of unknown length
      7131      7190: gap of unknown length
      7191      7250: gap of unknown length
      7251      7310: gap of unknown length
      7311      7370: gap of unknown length
      7371      7430: gap of unknown length
      7431      7490: gap of unknown length
      7491      7550: gap of unknown length
      7551      7610: gap of unknown length
      7611      7670: gap of unknown length
      7671      7730: gap of unknown length
      7731      7790: gap of unknown length
      7791      7850: gap of unknown length
      7851      7910: gap of unknown length
      7911      7970: gap of unknown length
      7971      8030: gap of unknown length
      8031      8090: gap of unknown length
      8091      8150: gap of unknown length
      8151      8210: gap of unknown length
      8211      8270: gap of unknown length
      8271      8330: gap of unknown length
      8331      8390: gap of unknown length
      8391      8450: gap of unknown length
      8451      8510: gap of unknown length
      8511      8570: gap of unknown length
      8571      8630: gap of unknown length
      8631      8690: gap of unknown length
      8691      8750: gap of unknown length
      8751      8810: gap of unknown length
      8811      8870: gap of unknown length
      8871      8930: gap of unknown length
      8931      8990: gap of unknown length
      8991      9050: gap of unknown length
      9051      9110: gap of unknown length
      9111      9170: gap of unknown length
      9171      9230: gap of unknown length
      9231      9290: gap of unknown length
      9291      9350: gap of unknown length
      9351      9410: gap of unknown length
      9411      9470: gap of unknown length
      9471      9530: gap of unknown length
      9531      9590: gap of unknown length
      9591      9650: gap of unknown length
      9651      9710: gap of unknown length
      9711      9770: gap of unknown length
      9771      9830: gap of unknown length
      9831      9890: gap of unknown length
      9891      9950: gap of unknown length
      9951      10000: gap of unknown length

```



Contact: leonowen@u.washington.edu  
 ----- Summary Statistics  
 Sequencing vector: pUC18: L08752  
 Chemistry: Dye-terminator Big Dye: 90% of reads  
 Chemistry: Dye-primer Big Dye: 10% of reads  
 Assembly program: Phrap: version 0.990399

NOTE: This record contains 96 individual  
 contigs that have been assembled into  
 contigs. The contigs of N are used to separate the reads  
 and the order in which they appear is completely  
 arbitrary. Low-pass sequence sampling is useful for  
 identifying clones that may be gene-rich and allows  
 overlap relationships among clones to be deduced.  
 However, it should not be assumed that this clone  
 will be sequenced to completion. In the event that  
 the record is updated, the accession number will  
 be preserved. 887: contig of 887 bp in length

1  
 888 1780: contig of 893 bp in length  
 1781 2645: contig of 865 bp in length  
 2646 3525: contig of 880 bp in length  
 3526 4401: contig of 876 bp in length  
 4402 5290: contig of 889 bp in length  
 5291 6155: contig of 865 bp in length  
 6156 7151: contig of 996 bp in length  
 7152 8036: contig of 885 bp in length  
 8037 8948: contig of 852 bp in length  
 8949 9836: contig of 888 bp in length  
 9837 10705: contig of 869 bp in length  
 10706 11579: contig of 874 bp in length  
 11580 12459: contig of 880 bp in length  
 12460 13321: contig of 852 bp in length  
 13322 14190: contig of 869 bp in length  
 14191 15058: contig of 868 bp in length  
 15059 15923: contig of 865 bp in length  
 15924 16802: contig of 879 bp in length  
 16803 17680: contig of 878 bp in length  
 17681 18548: contig of 868 bp in length  
 18549 19413: contig of 865 bp in length  
 19414 20289: contig of 876 bp in length  
 20290 21156: contig of 885 bp in length  
 21157 22031: contig of 875 bp in length  
 22032 22916: contig of 885 bp in length  
 22917 23786: contig of 870 bp in length  
 23787 24655: contig of 869 bp in length

24656 25519: contig of 864 bp in length  
 25520 26383: contig of 864 bp in length  
 26384 27253: contig of 870 bp in length  
 27254 28119: contig of 866 bp in length  
 28120 28987: contig of 868 bp in length  
 28988 29861: contig of 874 bp in length  
 29862 30737: contig of 876 bp in length  
 30738 31601: contig of 880 bp in length  
 31602 32481: contig of 880 bp in length  
 32482 33347: contig of 866 bp in length  
 33348 34218: contig of 871 bp in length  
 34219 35081: contig of 863 bp in length  
 35082 35948: contig of 867 bp in length  
 35949 36814: contig of 866 bp in length  
 36815 37687: contig of 873 bp in length  
 37688 38583: contig of 896 bp in length  
 38584 39452: contig of 869 bp in length  
 39453 40334: contig of 882 bp in length  
 40335 41201: contig of 867 bp in length  
 41202 42067: contig of 866 bp in length  
 42068 42933: contig of 866 bp in length  
 42934 43801: contig of 868 bp in length  
 43802 44675: contig of 874 bp in length  
 44676 45538: contig of 863 bp in length  
 45539 46405: contig of 867 bp in length  
 46406 47265: contig of 864 bp in length  
 47270 48139: contig of 880 bp in length  
 48140 48996: contig of 860 bp in length  
 49000 49866: contig of 867 bp in length  
 49867 50729: contig of 863 bp in length  
 50730 51588: contig of 859 bp in length  
 51589 52460: contig of 872 bp in length  
 52461 53323: contig of 863 bp in length  
 53324 54189: contig of 866 bp in length  
 54190 55053: contig of 864 bp in length  
 55054 55910: contig of 869 bp in length

5

[illegible]





|                 |                          |
|-----------------|--------------------------|
| repeat_c_region | /comp.fam5_12514a        |
| repeat_c_region | /rpf.family*~*1AL5C4     |
| repeat_c_region | 155_185                  |
| repeat_c_region | /rpf.family*~*ATGTCn     |
| repeat_c_region | complement(188_473)      |
| repeat_c_region | /rpf.family*~*AL5SG      |
| repeat_c_region | complement(14_1164C6)    |
| repeat_c_region | /rpf.family*~*AL5C4C6    |
| repeat_c_region | complement(168_1100)     |
| repeat_c_region | /rpf.family*~*AL5US*     |
| repeat_c_region | 1899_2178                |
| repeat_c_region | /rpf.family*~*AL5US      |
| repeat_c_region | complement(2635_2745)    |
| repeat_c_region | /rpf.family*~*12         |
| repeat_c_region | complement(2886_2897)    |
| repeat_c_region | /rpf.family*~*AL5U0      |
| repeat_c_region | 289_2936                 |
| repeat_c_region | /rpf.family*~*AL5U0C*    |
| repeat_c_region | complement(2937_3221)    |
| repeat_c_region | /rpf.family*~*AL5U0*     |
| repeat_c_region | 3871_3941                |
| repeat_c_region | /rpf.family*~*ALTR1P1    |
| repeat_c_region | 3964_4538                |
| repeat_c_region | /rpf.family*~*LM3C3*     |
| repeat_c_region | 4710_4951                |
| repeat_c_region | /rpf.family*~*HAL1*      |
| repeat_c_region | 4952_5040                |
| repeat_c_region | 5258_5554                |
| repeat_c_region | /rpf.family*~*HAL1*      |
| repeat_c_region | 5599_5900                |
| repeat_c_region | /rpf.family*~*AL5US*     |
| repeat_c_region | 5996_6158                |
| repeat_c_region | /rpf.family*~*AL5US*     |
| repeat_c_region | complement(6950_7202)    |
| repeat_c_region | /rpf.family*~*12*        |
| repeat_c_region | /rpf.family*~*AL5U0*     |
| repeat_c_region | complement(7509_7638)    |
| repeat_c_region | /rpf.family*~*12*        |
| repeat_c_region | complement(7639_8069)    |
| repeat_c_region | /rpf.family*~*M5TA*      |
| repeat_c_region | complement(8070_8120)    |
| repeat_c_region | /rpf.family*~*12*        |
| repeat_c_region | 8338_8666                |
| repeat_c_region | /rpf.family*~*AL4U*      |
| repeat_c_region | complement(9080_9282)    |
| repeat_c_region | /rpf.family*~*AL5US*     |
| repeat_c_region | complement(9456_9720)    |
| repeat_c_region | /rpf.family*~*ALTR1J*    |
| repeat_c_region | 9803_9967                |
| repeat_c_region | /rpf.family*~*AL4U*      |
| repeat_c_region | complement(11009_110392) |
| repeat_c_region | /rpf.family*~*AL5US*     |
| repeat_c_region | 10501_10557              |
| repeat_c_region | /rpf.family*~*AT_rich*   |
| repeat_c_region | 11080_11138              |
| repeat_c_region | complement(11167_11348)  |
| repeat_c_region | /rpf.family*~*M1R        |
| repeat_c_region | 11349_11708              |
| repeat_c_region | /rpf.family*~*M5E46*     |
| repeat_c_region | complement(11709_11758)  |
| repeat_c_region | /rpf.family*~*M1R        |
| repeat_c_region | 11784_11861              |
| repeat_c_region | /rpf.family*~*AL5US*     |
| repeat_c_region | complement(12131_12319)  |
| repeat_c_region | 12310_12339              |
| repeat_c_region | /rpf.family*~*TTCn*      |
| repeat_c_region | complement(12340_12457)  |
| repeat_c_region | /rpf.family*~*AL5U0*     |
| repeat_c_region | complement(12563_13020)  |

|               |                       |                            |
|---------------|-----------------------|----------------------------|
| repeat_region | /rpf_family_MuIR*     | comp1element(14090, 11428) |
| repeat_region | /rpf_family_M12*      | comp1element(15874, 15935) |
| repeat_region | /rpf_family_M59R      | comp1element(15972, 16163) |
| repeat_region | /rpf_family_M1625B    | comp1element(16256, 16366) |
| repeat_region | /rpf_family_M12*      | comp1element(18147, 18235) |
| repeat_region | /rpf_family_M14VNA    | 18331, 18509               |
| repeat_region | /rpf_family_MERSA*    | 18510, 18635               |
| repeat_region | /rpf_family_M1R3*     | comp1element(18811, 19056) |
| repeat_region | /rpf_family_M12*      | comp1element(19114, 19160) |
| repeat_region | /rpf_family_M12*      | 19794, 19947               |
| repeat_region | /rpf_family_M1R*      | comp1element(20054, 20111) |
| repeat_region | /rpf_family_MERSB     | 20112, 20420               |
| repeat_region | /rpf_family_M1AUSq    | 20797, 20968               |
| repeat_region | /rpf_family_M1AUSq    | comp1element(20715, 20798) |
| repeat_region | /rpf_family_MERSB     | 20799, 20968               |
| repeat_region | /rpf_family_M1AUSq/X* | 20972, 21283               |
| repeat_region | /rpf_family_M1AUSq    | comp1element(21287, 21327) |
| repeat_region | /rpf_family_MERSB     | 21357, 21795               |
| repeat_region | /rpf_family_M12*      | 21832, 21943               |
| repeat_region | /rpf_family_M1R*      | 21944, 22247               |
| repeat_region | /rpf_family_M1R*      | 22248, 22311               |
| repeat_region | /rpf_family_M1AUSq    | 22556, 22576               |
| repeat_region | /rpf_family_M1R_rich* | 23107, 23183               |
| repeat_region | /rpf_family_M1AUSq*   | 23655, 23976               |
| repeat_region | /rpf_family_M1AUSq*   |                            |

| Query Match | 4.98; | Score 232.2; | DB 9; | Length 172850; |
|-------------|-------|--------------|-------|----------------|
|-------------|-------|--------------|-------|----------------|

Matches 351; Conservative 0; Mismatches 53; Indels 37; Gaps 5;

QY 1 cagtgctgagcgagcagatctagagctgctgttcaggcttcagctgagtcgccgaac

Db 147267 CAGTGTGAGGCAGCAGGATGCAGAGTGTGTTCAAGTTTCCAGTGGGTCCTCAAG 147326

QY 61 ggga--ggaagaaacatctcta--ataacaaataagagagcttaacagtaac

Db 147327 GGACAACTTTGAAGATAGCTAGCTTCCAAGAATAAGAAATAGATAGAGTTACTGTAC

UY 115 Tgactctgggctctgccttaatccaagctgcttgcgcgcctgcaaggaaagataaatttcagaagcgt

DD 14/36/ 1GAC1GGGGC1GC1C11AA1CAGT1CTG-CA11GCAGGAAAG1AA1A11CGAGCG11 14/442

[illegible]

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2002, 14:52:41; Search time 206.96 Seconds

(Without alignment) 19593.875 Million cell updates/sec

Title: US-09-750-590-1

Perfect score: 4730  
1 cagctctgagcgagcgagat.....atcaaaaaaaaaaaaaaa 4730

Sequence: IDENTITY-MNC  
Gapop 10.0, Gapext 1.0

Scoring table: 930621 seqs, 428662619 residues

Searched: Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: N.Geneseq\_1101.\*  
2: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query | Match | Length | DB ID    | Description        |
|------------|--------|-------|-------|--------|----------|--------------------|
| 1          | 1752.8 | 37.1  | 2322  | 20     | AAV83134 | Polynucleotide clo |
| 2          | 1671   | 35.3  | 2222  | 22     | AAH75532 | Human CDNA sequenc |
| 3          | 1504.4 | 31.8  | 1980  | 20     | AAH40197 | Sequence of C17A7  |
| 4          | 1215.6 | 25.7  | 1628  | 22     | AAH13782 | Human CDNA sequenc |
| 5          | 600.2  | 12.7  | 848   | 22     | AAH03594 | Human CDNA clone ( |
| 6          | 460.2  | 9.7   | 571   | 21     | AAA45014 | Human secreted exp |
| 7          | 342.4  | 7.2   | 441   | 22     | AA129101 | Colon tumour relat |
| 8          | 296.4  | 6.2   | 371   | 22     | AAH66083 | Novel human polynu |
| 9          | 279.6  | 5.9   | 407   | 20     | AAH68305 | EST clone D0410.   |
| 10         | 259.6  | 5.3   | 401   | 22     | AAH66074 | Novel human polynu |
| 11         | 230    | 4.9   | 421   | 21     | AAH77915 | CDNA encoding huma |

| Result No. | Score | Query | Match | Length | DB ID    | Description        |
|------------|-------|-------|-------|--------|----------|--------------------|
| 12         | 230   | 4.9   | 421   | 22     | AA128653 | Colon tumour relat |
| 13         | 230   | 4.9   | 421   | 22     | AA129076 | Colon tumour relat |
| 14         | 221.6 | 4.7   | 1434  | 22     | AAH22225 | Human sa.79 gene.  |
| 15         | 210   | 4.4   | 361   | 21     | AAH27789 | Human secreted pro |
| 16         | 199.8 | 4.2   | 279   | 21     | AAA11967 | Human secreted exp |
| 17         | 147   | 3.1   | 2740  | 22     | AAH16256 | Human CDNA sequenc |
| 18         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 19         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 20         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 21         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 22         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 23         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 24         | 144.6 | 3.1   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 25         | 143.4 | 3.0   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 26         | 143.4 | 3.0   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 27         | 143.4 | 3.0   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 28         | 143.4 | 3.0   | 936   | 22     | AAH58232 | Oligonucleotide D1 |
| 29         | 137.5 | 2.9   | 2407  | 20     | AAH24392 | Oligonucleotide D1 |
| 30         | 137.5 | 2.9   | 2407  | 20     | AAH24392 | Oligonucleotide D1 |
| 31         | 131   | 2.8   | 1712  | 22     | AA152027 | Human polynucleoti |
| 32         | 87.6  | 1.9   | 3399  | 17     | AAH03688 | Chicken leucocytos |
| 33         | 85.2  | 1.8   | 6644  | 20     | AAH33181 | Base sequence of t |
| 34         | 85.2  | 1.8   | 7372  | 20     | AAH33182 | Base sequence of t |
| 35         | 85.2  | 1.8   | 7797  | 20     | AAH33180 | Base sequence of t |
| 36         | 85.2  | 1.8   | 7996  | 20     | AAH33184 | Base sequence of t |
| 37         | 76.9  | 1.7   | 3579  | 21     | AAH40099 | Human CDNA clone ( |
| 38         | 76.9  | 1.7   | 3579  | 21     | AAH40099 | Human CDNA clone ( |
| 39         | 75.2  | 1.6   | 1686  | 12     | AAH87529 | Plasmodium falcipa |
| 40         | 73    | 1.5   | 5940  | 21     | AAH70105 | Plasmodium falcipa |
| 41         | 71.4  | 1.5   | 653   | 22     | AAH33568 | Human colon cancer |
| 42         | 70.8  | 1.5   | 1998  | 21     | AAH70212 | Plasmodium falcipa |
| 43         | 68.2  | 1.4   | 355   | 21     | AAH27957 | Human secreted pro |
| 44         | 68.2  | 1.4   | 356   | 20     | AAH41086 | Human secreted pro |
| 45         | 67.4  | 1.4   | 1797  | 22     | AAH90586 | Degenerate DNA cod |

## ALIGNMENTS

| Result No. | Score  | Query | Match | Length | DB ID    | Description        |
|------------|--------|-------|-------|--------|----------|--------------------|
| 1          | 1752.8 | 37.1  | 2322  | 20     | AAV83134 | Polynucleotide clo |
| 2          | 1671   | 35.3  | 2222  | 22     | AAH75532 | Human CDNA sequenc |
| 3          | 1504.4 | 31.8  | 1980  | 20     | AAH40197 | Sequence of C17A7  |
| 4          | 1215.6 | 25.7  | 1628  | 22     | AAH13782 | Human CDNA sequenc |
| 5          | 600.2  | 12.7  | 848   | 22     | AAH03594 | Human CDNA clone ( |
| 6          | 460.2  | 9.7   | 571   | 21     | AAA45014 | Human secreted exp |
| 7          | 342.4  | 7.2   | 441   | 22     | AA129101 | Colon tumour relat |
| 8          | 296.4  | 6.2   | 371   | 22     | AAH66083 | Novel human polynu |
| 9          | 279.6  | 5.9   | 407   | 20     | AAH68305 | EST clone D0410.   |
| 10         | 259.6  | 5.3   | 401   | 22     | AAH66074 | Novel human polynu |
| 11         | 230    | 4.9   | 421   | 21     | AAH77915 | CDNA encoding huma |





```

Db      1741  atctgcaaaagatgagaaagatctgcttcttcagatcttgagcaagaatctaaagatca 1800
      |||||
Oy      4207  gcaaggacgatgtgacaaatccttaacacacacacagagctacagaaagaatcacagga 4266
      |||||
Db      1801  gaagaaacgatgtgacaaagctcttaacaaacacacagagctacagaaagaatcacagga 1860
Oy      4267  atctgcaaaacaaatcgaaagcaaaaataaataagataactgacatgcttccatgatatgga 4326
      |||||
Db      1861  atctgcaaaacaaatcgaaagcaaaaataaataagataactgacatgcttccatgatatgga 1920
Oy      4327  gagattaaacagagccctcaatgtgcttctccagctacacttaatgaaatggagagccagg 4386
      |||||
Oy      1921  aagattaaacagagccctcaatgtgcttctccagctacacttaatgaaatggagagccagg 1980
      |||||
Oy      4387  caaaggagaaagatcagaatctatcttctcagctcagagcaggttgaagctccctcagagaga 4446
      |||||
Db      1981  caaaggagaaagcagagctgattgagacactcttctcagagcaggaagtgaaatctcttgagacaag 2040
Oy      4447  gctgagagatgctgcagacagacagcaagaagaatgtcaattatctcgagacacactctct 4506
      |||||
Db      2041  gctgagcgaatgctgcagacagacacagaaagaatgtcaattatctcgagacacacactctct 2100
Oy      4507  taagtgcacagagatccacatggaatgagatgtgacagccagctactctgacagatcacaa 4566
      |||||
Db      2101  taagtgcacagagatccacatggaatgagatgtgacagccagctactctgacagatcacaa 2160
Oy      4567  gatgagcgaagagatcgtgctgctgctatgagc--gacacccacagacacagagcttccct 4623
      |||||
Db      2161  aatcgagcaagagcgtgctgctgctatgagccttgaagcagctgacagatcgtttatctt 2220
Oy      4624  gctggtgctgagcatctctgagcaacttcagtgcttcttggagcctcgctgctgctagaa 4683
      |||||
Db      2221  gctggtgctgagcatctctgagcaacttcagtgcttcttggagcctcgctgctgctagaa 2280
Oy      4684  taacttaataataagatgattttatcttcaatcaaaaataaaaaaa 4725
      |||||
Db      2281  taacttaataataataatatttcttcttaaaaaaaataaaaaaa 2322
      |||||

```

```

DR      WPI: 2001-483899/53.
P-PSDB: A064560.
XX      Polypeptide-human myosin heavy chain 74 and polynucleotide for coding
PT      said polypeptide -
XX      Claim 6; Page 25-26 Disclosure; 36pp; Chinese.
XX      The invention relates to human myosin heavy chain 74 useful for treating
CC      diseases, e.g. cancer, human immunodeficiency virus (HIV) infection. The
CC      present sequence is that of the human myosin heavy chain 74 encoding
CC      cDNA.
XX      Sequence 2222 BP; 905 A; 351 C; 479 G; 487 T; 0 other;
SQ
Query Match      35.38; Score 1671; DB 22; Length 2222;
Best Local Similarity 85.2%; 2264 N/A; 0;
Matches 1892; Conservative 0; Mismatches 320; Indels 9; Gaps 2;
Oy      2495  aagaaatcgaaaagatgctgcctgagataaagcttcccttacaacaaagataaatactta 2554
      |||||
Db      1  aagaaatcgaaaagatgctgcctgagataaagcttcccttacaacaaagataaatactta 60
Oy      2555  acaatgaaatgaaaa-----atgtcctcttaaaagttaagtgaagaatgtaaaaagta 2608
      |||||
Db      61  caaatgaaatgaaaatactcaatcagatctcttcaaaagtaagtgaaagacagaaagatca 120
Oy      2609  catgattgaatgtgtgattgattgataaagaagcttccagagatgagacacaaatataca 2668
      |||||
Db      121  catgattgaatattgattgattccttaagaagcttttagattgtaacacaaatataca 180
Oy      2669  gaaaagaagtctgaaaatcgaaagatgtgctactgtaaaaaagccagtttaagataaaatgctc 2728
      |||||
Db      181  gaaaagaagtctgaaaatcgaaagatgtgctactgtaaaaaagccagtttaagataaaatgctc 2788
      |||||
Oy      2729  agccgcctgaaaatcgtgttcaatcacttccagagacagaaagaagaatgtgtgtctg 2788
      |||||
Db      241  agccgcctgaaaatcgtgttctactctcctgagaaataatgaaagaagataatgtgtctg 300
Oy      2789  aatccaatatcactgaaacttaagaagcagctgctgaaacttaataaaaaatgtgtgtga 2848
      |||||
Db      301  aatccaatatcactgaaacttaagaagcagctgctgaaacttaataaaaaatgtgtgtga 3060
Oy      2849  gaccaaagaaaataatattcactcaatgctgcgaaaacaaatgatttgaaaaagccatgagt 3208
      |||||
Db      361  gaccaaagaaaataatattcactcaatgctgcgaaaacaaatgatttgaaaaagccatgagt 420
Oy      2909  catcagatgtgtccgtgaaaacccacatgaaaggaatctaaacgtctgagtgtgacattg 2968
      |||||
Db      421  aatcagatgtgtccgtgaaaacccacatgaaaggaatctaaacgtctgagtgtgacattg 480
Oy      2965  gataaaacaaatcagaatctgattgattgtaagaagaatgtctaaagataataatacaaga 3028
      |||||
Db      481  gcaaaacaaatcagaatctgattgattgtaagaagaatgtctaaagataataatacaaga 540
Oy      3029  ttgtgaaaataaagaatgaaagaaatcttaaaagaagaatcttggaaacacctcgagac 3088
      |||||
Db      541  ttgtgaaaataaagaatgaaagaaatcttaaaagaagaatcttggaaacacctcgagac 600
Oy      3089  caagttaaagctgagtcacgtcagctcgaagagcagatgaaagaagaatgaaatgagctaaag 3148
      |||||
Db      601  caaataaagctcagatcatcagctcgttcgagagcagagaaagatgaaatgagctaaag 650
Oy      3149  aagaatctgaaagaatgtccagaagacagagctgaaatctctgacaaagtacaaaaagac 3208
      |||||
Db      661  aagaatctgaaagaatgtcagagatgtaataatgattgattgcttcgcaactctcgaaaaagc 720
Oy      3209  caggagagattgtcacaccctgcatgagaaatctgacagccagaaagaaagaaactcgacag 3268
      |||||
Db      721  caagaagaatgtgtgacatcgtcagaaatctaaagccagaaagaaagaaactcgacagaa 780
Oy      3265  atacagaatgatacaagtaataataatgctccgatacatcagctcttgaagaagtgtgaagaga 3328

```



QY 1127 tccaacaaggaggaactctggaagaaaggaaccatcttacaacaagcgaatatctgtct 1

[illegible]













|    |                                                                    |                         |
|----|--------------------------------------------------------------------|-------------------------|
| ID | AA666074                                                           | standard: cDNA: 401 bp. |
| XX | AA666074:                                                          |                         |
| AC |                                                                    |                         |
| XX |                                                                    |                         |
| DT | 09-APR-2001                                                        | (first entry)           |
| XX |                                                                    |                         |
| DE | Novel human polynucleotide, SEQ ID NO: 1830.                       |                         |
| KX | Human: cytosolic; gene therapy; colon cancer; prostate cancer;     |                         |
| KM | breast cancer; lung cancer; cancer detection; ss.                  |                         |
| XX |                                                                    |                         |
| OS | Homo sapiens.                                                      |                         |
| XX |                                                                    |                         |
| PN | W0200102556-A2.                                                    |                         |
| XX |                                                                    |                         |
| PI | 11-JAN-2001.                                                       |                         |
| XX |                                                                    |                         |
| PE | 30-JUN-2000: 2000MO-US18374.                                       |                         |
| XX |                                                                    |                         |
| PR | 02-JUL-1999: 99US-0142310.                                         |                         |
| PR | 02-JUL-1999: 99US-0142311.                                         |                         |
| XX |                                                                    |                         |
| XX | (CHIR ) CHIRON CORP.                                               |                         |
| PA | (HIS-) HISCO INC.                                                  |                         |
| XX |                                                                    |                         |
| PI | Williams LW, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A, |                         |
| PI | Reinhard C, Randoz F, Kennedy GC, Pot D, Lamson G, Dameson R;      |                         |
| PI | Chernyshev G, Dimahe S, Dickson M, Labat I, Leshkowitz D;          |                         |
| PI | Kita D, Garcia V, Jones LW, Strache-Crain B;                       |                         |
| XX |                                                                    |                         |
| XX | WPI: 2001-091805-10.                                               |                         |

**Claim 9:** Page 805; 106dpp; English.

The present sequence is one of 3351 sequences in a library of human polynucleotides. The library is used to detect differentially expressed genes correlated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional libraries or probe sets for detection of specific mRNAs associated with carcinomas or antisense oligonucleotides can be generated. The polynucleotides are blood or tissue products that will detect the earliest changes along the carcinogenesis pathway and/or monitor the efficacy of therapies and preventative interventions. The polynucleotides, peptides and nucleosides treat the cancers and proliferative disorders such as neoplasia, dysplasia and hyperplasia.

Sequence 401 BP; 114 A; 85 G; 107 G; 93 T; 2 other;

|                           |        |                    |           |             |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match               | 5.50;  | Score 259.8;       | DB 22;    | Length 401; |
| Best Local Similarity     | 87.80; | Pred. No. 2.6e-52; |           |             |
| Matches 294; Conservative | 0;     | Mismatches 39;     | Indels 2; | Gaps 1;     |

|    |     |                                                                       |     |
|----|-----|-----------------------------------------------------------------------|-----|
| Oy | 726 | gtccaccatgaacatgtagacccctccgggaaggaactgcgaattctatgtctcagctctaaagccgg  | 785 |
| Db | 726 | gtccaccatgagatgagatgacccctccgggaaggaactgcgaattctatgtctcagctctaaagccgg | 845 |
| Oy | 786 | actgctctctatgataatcagctgtccctcggagaaacacgtcctcgtatgacgaacatggt        | 845 |
| Db | 786 | actgctctctatgataatcagctgtccctcggagaaacacgtcctcgtatgacgaacatggt        | 905 |
| Oy | 846 | taagatggcgcgacacacatgcttcctcggtaccacagatgctgtagcacaacataatgataaac     | 905 |
| Db | 846 | taagatggcgcgacacacatgcttcctcggtaccacagatgctgtagcacaacataatgataaac     | 965 |

QY 506 tggatagataaagaggggagatataatctcgaagaaacaaatgctgcctca 963  
 Db 208 tgcgtgaatagataagagagagatgaatctccgaagaaacaaatgagctccca 267  
 QY 966 tgcgtgaatagagagatgagtcgaatgaagagagagcttaataaaagcggctg 1025  
 Db 966 tgcgtgaatagagagatgagtcgaatgaagagagagcttaataaaagcggctg 1025  
 QY 288 tgcgtgagttgagaaatggtctgcagagatgcacgaataaagcttaaat~aaatcgtgctg 325  
 Db 1026 acgtgacctctcgagcgcctctgtgcacaaagcaat 1060  
 QY 326 atataagctgcgtgcgtgcctcagaaagaaatg 360

RESULT 11  
AAA77915  
ID AAA77915 standard; cDNA; 421 BP.

DT 14-NOV-2000 (first entry)

DE CDNA encoding human colon tumour protein CT607, SEQ ID NO:195

KM Human colon tumour polypeptide; tumour antigen; cancer; vaccine  
KW Immunotherapy; diagnosis; progression; ss.

**Homo sapiens.**

PN W02000037643-A2.

PD 29-JUN-2000

PF 23-DEC-1999; 99WO-US30909.

PR 23-DEC-1998; 98US-0221298.

PR 22-SEP-1999; 99US-0401064.

PR 02-DEC-1999; 99US-0454150.

XX  
PA  
(CORI-) CORIXA CORP.

|    |      |          |             |           |            |           |
|----|------|----------|-------------|-----------|------------|-----------|
| XX | yu T | Iodas MI | Secretist H | Benson DB | Meagher MT | Strojk JT |
| XX | BT   |          |             |           |            |           |

PI Wang T, Yuqiu J;

DR WPI; 2000-442671/38

XX

especially colon cancer, and for diagnosing and monitoring the

XX  
progression of the cancer -

Claim 29; page 14 /; 229pp; English.

CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or

CC The invention also specifically discloses a human colon tumour protein

CC antigen presenting cells (APCs, preferably dendritic cells) expres

especially colon tumour cells, thereby inhibiting the development

used to remove tumour cells from biological samples, especially bl

polypeptide can then be used to inhibit cancer development. CD4+ a

CC CB8+T-cells from a patient may be incubated with a polypeptide or

CC to cause the proliferation of specific T-cells. The T-cells can be collected and then administered back to the patient to inhibit cancer.

development. Nucleic acids encoding the polypeptides and antibodies

CC of a tumour protein of the invention, and therefore to determine whether

CC monitor the progression of a cancer by repeating the process at time  
 CC intervals, and comparing the current result to previous results. The  
 CC present sequence represents a cDNA encoding a human colon tumour  
 CC polypeptide.  
 CC XX

50 Sequence 421 BP, 109 A, 101 C, 125 G, 86 T, 0 other;

Query Match 4.9%: Score 230; DB 21; Length 421;  
 Best Local Similarity 92.4%; Pred. No. 3, 3e-45;

Matches 242; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

431 caagcagcagatctggaacaaatacagatcgtatgagagcagcagaagagggagat 490  
 160 cagcagcagcagatctggaacaaatacagatcgtatgagagcagcagaagagggagat 219  
 491 gtagaagaaagtgctcctatcctgtctaaagagatcgaacagagcagcagatc 550  
 220 gtagaagaaagtgctcctatcctgtctaaagagatcgaacagagcagcagatc 279  
 551 gaagcagcagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 610  
 280 gaagcagcagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 339  
 611 atcctatacagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 670  
 340 atcctatacagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 399  
 671 ctgagctcagaagtagggcagc 692  
 400 ctgagctcagaagtagggcagc 421

## RESULT 12

AA128653  
 ID AA128653 standard; cDNA: 421 BP.

AA128653:

12-OCT-2001 (first entry)

DE Colon tumour related determined cDNA sequence for CT607.

KW Human: immunotherapy; diagnosis; colon cancer; colon tumour;

IM Immunogenic; gene therapy; vaccine; colonic cancer; ss.

XX Homo sapiens.

XX MO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000US-0635596.

XX 30-DEC-1999; 99US-0476296.

XX 10-JAN-2000; 2000US-0480321.

XX 06-FEB-2000; 2000US-0504629.

XX 19-MAY-2000; 2000US-0575251.

XX 28-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xn J, Lodes M, Secretat H, Henson DR, Meagher MJ, Stolk JA;

XX King GE, Wang T, Jiang Y;

XX WPI: 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the

XX prevention, diagnosis and treatment of colonic cancer.

XX Claim 25: Page 187: 472pp; English.

XX The present invention describes colon tumour associated proteins (i) and  
 CC (ii) that encode them. (i) have cytoskeletal activity,  
 CC (i) and (ii) can be used in the prevention, diagnosis and treatment of  
 CC (ii) may be used in the prevention, diagnosis and treatment of  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (i) and (ii) may be  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC produce the TCAP proteins. Additionally, (ii) may be used to  
 CC cell culturing the cell to express the protein. (i) and (ii)  
 CC complementary sequences may also be used as DNA probes to detect and  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (i) may  
 CC also be used as antigenic modulators of TCAP expression and activity.  
 CC Anti-(i) and (ii) antibodies may be used to down regulate  
 CC TCAP expression and activity. The anti-(i) antibodies may be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128650 to AA129512  
 CC given in the exemplification of the present invention.

50 Sequence 421 BP, 109 A, 101 C, 125 G, 86 T, 0 other;  
 Query Match 4.9%: Score 230; DB 22; Length 421;  
 Best Local Similarity 92.4%; Pred. No. 3, 3e-45;

Matches 242; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

431 caagcagcagatctggaacaaatacagatcgtatgagagcagcagaagagggagat 490  
 160 cagcagcagcagatctggaacaaatacagatcgtatgagagcagcagaagagggagat 219  
 491 gtagaagaaagtgctcctatcctgtctaaagagatcgaacagagcagcagatc 550  
 220 gtagaagaaagtgctcctatcctgtctaaagagatcgaacagagcagcagatc 279  
 551 gaagcagcagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 610  
 280 gaagcagcagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 339  
 611 atcctatacagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 670  
 340 atcctatacagatctgctcttccttcctgctgagcctcaagaggaactctgagtgagc 399  
 671 ctgagctcagaagtagggcagc 692  
 400 ctgagctcagaagtagggcagc 421

## RESULT 13

AA129076  
 ID AA129076 standard; cDNA: 421 BP.

AA129076:

12-OCT-2001 (first entry)

DE Colon tumour related determined cDNA sequence for clone CT607.

KW Human: immunotherapy; diagnosis; colon cancer; colon tumour;

IM Immunogenic; gene therapy; vaccine; colonic cancer; ss.

XX Homo sapiens.

XX MO200149716-A2.

XX 12-JUL-2001.



Query Match 4.7% Score 221.6; DB 22; Length 1434;  
 Best Local Similarity 52.7%; Prod. No. 6.2e-43;  
 Matches 395; Conservative 0; Mismatches 291; Indels 0; Gaps 0;

Oy 422 aagaaagacacagacagatctggaacaaatcacagacatctgattgagagacagaa 481  
 Db 153 aggaagagtgatccatagagatgagaaagaaatgagatcgatctgacagagcagag 212  
 Oy 482 agggagaaagagagagagagagagagagagagagagagagagagagagagag 541  
 Db 213 aatggaaabagagagagagagagagagagagagagagagagagagagagagag 272  
 Oy 542 ctatgagctgaaagagagatcgatctgagctgagctgagctgagagagagagag 601  
 Db 273 cccagacagagagagagagagagagagagagagagagagagagagagagag 332  
 Oy 602 tggagctgacatctgacatctgagatgagatctgacacagagagagagagagat 661  
 Db 333 ctcagagagatgatatcacatctgagctgagctgagctgagctgagctgagctg 392  
 Oy 662 gctcttcaacctgctgagagatgagagatgagctgagctgagagagagagagag 721  
 Db 393 gctcttcaacctgagagagagagagagagagagagagagagagagagagagagag 452  
 Oy 722 aatgctgacatctgagagagagagagagagagagagagagagagagagagagag 781  
 Db 453 aatgctgacagagagagagagagagagagagagagagagagagagagagagag 512  
 Oy 782 gacagctgctctctgagatcacagctgagctgagagagagagagagagagag 841  
 Db 513 cagagctgagctctctgagatcacagctgagagagagagagagagagagagag 572  
 Oy 842 gatttgatgagagagagagagagagagagagagagagagagagagagagagag 901  
 Db 573 gatttgatgagagagagagagagagagagagagagagagagagagagagagag 632  
 Oy 902 caactctgagatcacagagagagagagagagagagagagagagagagagagagag 961  
 Db 633 caactctgagatcacagagagagagagagagagagagagagagagagagagagag 692  
 Oy 962 ctctctgagatcacagagagagagagagagagagagagagagagagagagagag 1021  
 Db 693 ctctctgagatcacagagagagagagagagagagagagagagagagagagagag 752  
 Oy 1022 gctctgagatcacagagagagagagagagagagagagagagagagagagagag 1081  
 Db 753 gctctgagatcacagagagagagagagagagagagagagagagagagagagag 812  
 Oy 1082 gctctgagatcacagagagagagagagagagagagagagagagagagagagag 1107  
 Db 813 gctctgagatcacagagagagagagagagagagagagagagagagagagagag 838

RESULT 15  
 AAC27789  
 ID AAC27789 standard; CDNA; 361 BP.  
 AC  
 AC AAC27789;  
 DT 06-OCT-2000 (first entry)  
 XX  
 XX Human secreted protein 5' EST, SEQ ID NO: 31864.  
 DE Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;  
 KM gene therapy; chromosome mapping; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 PN EPI033401-A2.  
 XX  
 XX 06-SEP-2000.  
 PD  
 XX 21-FEB-2000; 2000BP-0200610.

XX  
 XX 26-FEB-1999; 990S-0122487.  
 XX  
 XX (GEST) GENSET.  
 XX  
 XX Dumas, Mline Edwards J., Duclert A., Giordano J;  
 DR  
 DR WPI: 2000-500381/45.  
 XX  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 XX Claim 1: SEQ ID 31864; 71bp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC human poly(A) mRNAs derived from 30 different tissues. EST  
 CC sequences usually correspond to a region of the mRNA that is located  
 CC 11bp from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC for diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC The present sequence is a human regulatory sequences and to design  
 CC expression and secretion vectors.  
 CC  
 CC Sequence 361 BP; 112 A; 75 C; 89 G; 83 T; 2 other;

Query Match 4.4% Score 210; DB 21; Length 361;  
 Best Local Similarity 92.7%; Prod. No. 1.8e-40;  
 Matches 230; Conservative 1; Mismatches 16; Indels 1; Gaps 1;

Oy 709 actctcagatgacatctgctcctcagatcacagctgagagagagagagagagagag 768  
 Db 20 actctcagatgacatctgctcctcagatcacagctgagagagagagagagagagag 79  
 Oy 769 agatgagctatgagacatctgctcctcagatcacagctgagagagagagagagag 828  
 Db 80 agatgagctatgagacatctgctcctcagatcacagctgagagagagagagagag 139  
 Oy 829 agatgagctatgagacatctgctcctcagatcacagctgagagagagagagagag 887  
 Db 140 agatgagctatgagacatctgctcctcagatcacagctgagagagagagagagag 199  
 Oy 888 ggcctcaacatctgacatctgctcctcagatcacagctgagagagagagagagag 947  
 Db 200 ggcctcaacatctgacatctgctcctcagatcacagctgagagagagagagagag 259  
 Oy 948 aaacagag 955  
 Db 260 aaacagag 267

Search completed: April 10, 2002, 15:57:16  
 Job time: 3875 sec

Thu Apr 11 07:43:27 2002

us-09-750-590-1.std.rng

---

Page 15

**THIS PAGE BLANK (USPTO)**



```

Query Match Summary: 2.2% Score 103.2; Da 1; Length 7218;
Overall Similarity: 9.8%; Pred. No. 5,4e-16;
Matches: 45; Conservatave: 255; Mismatches: 158; Indels: 0; Gaps: 0

Db 1764 AATGCAAGGCGATGATCATTGAAATGAGAGAGGATCTCGATGAGCAAGAA 1823
Oy 1824 AGCAACCAAGAGATGAGCTCTTAAGAGCGAGAGAGAGATGATGGTACRRRRRRR 1427
Db 1486 ACTGATATACCTATCTATGACAGCTTATTAAGCATGTAAGCAATTGGTACRRRRRR 1427
Oy 1824 AGCAACCAAGAGATGAGCTCTTAAGAGCGAGAGAGAGATGATGGTACRRRRRRR 1483
Db 1426 RRR 1367
Oy 1884 TGAAACAATGAGCAACAATCTTCTGCTTGAAGAGCACTGACAAATGATGAGCGCA 1943
Db 1366 RRR 1307
Oy 1944 CGGAGACCAACGAGCTGATGAGAGACTGAAGATCGATGGAAGCATGAAGAGAAAT 2003
Db 1306 RRR 1247
Oy 2004 AGGAGAGTGGCTGCGAGAGTGGGAGAAATGGAACAACAACAAGAGAAATGCT 2063
Db 1246 RRR 1187
Oy 2064 TGGTGTGAGAGTGAAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2123
Db 1186 RRR 1187
Oy 2124 AGAGGTATGAGCATGCGTGAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2193
Db 1126 RRR 1067
Oy 2184 AAGGCAAGTGAAGAGATCTGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2221
Db 1066 ATCCAGACATCTCCCTGCAGCTGCAGCTGCAGCTGCAGATT 1029

RESULT 2
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Candelo, John J.
; APPLICANT: Candelo, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, William, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Kansas
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Windows 3.11.0, Version #1.30
; CURRENT APPLICATION DATES: 08/23/99
; APPLICATION NUMBER: US/08/931.999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Callahan, William H.
; REGISTRATION NUMBER: 36,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:

```

[illegible]





DB 944 ARCGNCTGACROTNMNSAAYARARARABRGNMAARTTNCNWSNGAYGTNMAAYGARGNA 1003  
 OY 3474 aggaabaagacgtccacatlgaggaactctatgaaacagaagagcalttaagccagaanaa 3533  
 DB 1004 ARACGNTTATTHGNCNAAYTNMSTTTGATWNSCARCARCARATNGCNGARTNY 1063  
 OY 3534 cagaagagcctgaaacagcttaaaagacgctgtgcagaatccccaagagccagaag 3593  
 DB 1064 TNCACARTTTCGNCARTYNAARTYATGTMGNATHGTYTNCAYCCNGAYACNACAYW 1123  
 OY 3594 agaaag 3599  
 DB 1124 SNAAGC 1129

RESULT 5  
 US-08-676-974-2  
 Sequence 2, Application US/08676974  
 Request No. 57442  
 GENERAL INFORMATION:  
 APPLICANT: COLLINS, KATHLEEN  
 TITLE OF INVENTION: Human Telomerase  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Science & Technology Law Group  
 STREET: 268 Bush Street, Suite 3200  
 STATE: San Francisco  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/676, 974  
 FILING DATE:  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Osman Ph.D., Richard A  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: UCS96-055  
 TELEPHONE: (415)343-4341  
 TELEFAX: (415)343-4342  
 INFORMATION FOR SPO ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2277 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 Molecule type: cDNA  
 US-08-676-974-2

Query Match 1.2% Score 54.6 DB 1 Length 2277  
 Best Local Similarity 25.5% Pzed. No. 0.00043  
 Matches 216: Conservative 134; Mismatches 433; Indels 3; Gaps 1

DB 464 GNTTTCGNTTGTTCARATTAAATTTTATTCRCGNCGNAAACRCYTNAAARCGNATGA 523  
 OY 2997 tgaagaagaagagtgagaagataataatcaagaatttgaaataaagaatgagacgaa 3056  
 DB 534 AYATGACRARRATTHAARGGNGNACNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593  
 OY 3057 taataaagaagaaacccggaagaaacacacacacacacacacacacacacacacacacac 3116  
 DB 584 ATAAAGATACGNCARMSNTTMSNCGNATHGNCGRARARMSNCAVGARMSNAAACAYC 643  
 OY 3117 ggaagctgaagaagaagtgatgtgctgaagaagacatgaagaagccagagacaca 3176  
 DB 644 AGARMSNCTNAARARARAGGNGNACGNCARCARCARATTCGNCARARARARARARAR 703  
 OY 3177 ggcctgaaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 3236  
 DB 704 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 763  
 OY 3237 ag--atgagcccaagagagagagagagagagagagagagagagagagagagagagagag 3293  
 DB 764 ABAATATHARMSNAARCTGACNARACNCTGNCARTTTCARARARARARARARARARAR 823  
 OY 3294 agctcgcgcacacacgctctggaagagtgagagagagagagagagagagagagagagagag 3353  
 DB 824 CAGCNCNCNCAARMSNMSNGAYCAVYMSNGARCARAVMSNCAVYTGARCARMSNGAYW 883  
 OY 3354 taagaagaagcctaccccgagcagcagagagatataacagtgagaagagagccaga 3413  
 DB 884 SMATHGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 943  
 OY 3414 agtgcagaagaagaatgaaagcttaagaagaagagagagagagagagagagagagagagag 3473  
 DB 944 ARCGNCTGACROTNMNSAAYARARARABRGNMAARTTNCNWSNGAYGTNMAAYGARGNA 1003  
 OY 3474 aggaabaagacgtccacatlgaggaactctatgaaacagaagagcalttaagccagaanaa 3533  
 DB 1004 ARACGNTTATTHGNCNAAYTNMSTTTGATWNSCARCARCARATNGCNGARTNY 1063  
 OY 3534 cagaagagcctgaaacagcttaaaagacgctgtgcagaatccccaagagccagaag 3593  
 DB 1064 TNCACARTTTCGNCARTYNAARTYATGTMGNATHGTYTNCAYCCNGAYACNACAYW 1123  
 OY 3594 agaaag 3599  
 DB 1124 SNAAGC 1129

RESULT 6  
 US-09-098-487-2  
 Sequence 2, Application US/09098487  
 Request No. 57442  
 GENERAL INFORMATION:  
 APPLICANT: COLLINS, Kathleen  
 TITLE OF INVENTION: Human Telomerase  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Science & Technology Law Group  
 STREET: 268 Bush Street, Suite 3200  
 STATE: San Francisco  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/098,487  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:



OY 967 gctggatgcgagatcgtgcacaaagtcgtagaagcttcaacaaacagcgcgc 1025  
 Db 365 catggacaacggatcgtgcacaaagctgcacacacttcaatcaccagctgagc 423

## RESULT 8

US-08-631-097-3  
 : Sequence 3, Application US/08631097  
 : Patent No. 5968816  
 : GENERAL INFORMATION:  
 : TITLE OF INVENTION: Adenovirus Vector System, Adenovirus  
 : TITLE OF INVENTION: Adenovirus Vector System, Adenovirus  
 : NUMBER OF SEQUENCES: 7  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Migma, Cohen, Leitner, & Myers, P.C.  
 : STREET: 900 17th Street, N.W., Suite 1000  
 : CITY: Washington  
 : STATE: D.C.  
 : ZIP: 20006  
 : COUNTRY: USA  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 Mb storage  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: DOS  
 : SOFTWARE: ASCII  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/631,097  
 : CLASSIFICATION: 12 Apr-96  
 : FILING DATE: 12-Oct-94  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: PCT/US94/11598  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Cohen, Herbert  
 : REGISTRATION NUMBER: 25,109  
 : REFERENCE/DOCKET NUMBER: 0744,057  
 : TELEPHONE: (202)471-6915  
 : TELEFAX: (202)471-6915  
 : INFORMATION FOR SEQ ID NO: 3:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 4935 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: Linear  
 : NO. OF FEATURES: 1  
 : HYPOTHEICAL: NO  
 : ANTI-SENSE: NO  
 : FRAGMENT TYPE: No. 5968816 applicable  
 : ORIGINAL SOURCE:  
 : ORGANISM: homo sapiens  
 : STRAIN: not applicable  
 : DEVELOPMENTAL STAGE: not applicable  
 : TISSUE TYPE: not applicable  
 : CELL LINE: HeLa  
 : ORGANELLAE: not applicable  
 : IMMEDIATE SOURCE:  
 : LIBRARY: not applicable  
 : CLONE: not applicable  
 : POSITION IN GENOME:  
 : CHROMOSOME/SEGMENT: not applicable  
 : MAP POSITION: not applicable  
 : UNITS: not applicable  
 : FEATURE:  
 : NAME/KEY: Seq. ID. NO.: 3 1s  
 : NAME/KEY: the sequence in claim 1(11) as Figure 8 of the specification  
 : LOCATION: not available  
 : IDENTIFICATION METHOD: experiment-  
 : IDENTIFICATION METHOD: In specification

OTHER INFORMATION: prevention of IFN-2  
 : OTHER INFORMATION: promoted cell death  
 : PUBLICATION INFORMATION: not available  
 : US-08-631-097-3

## Query Match

11% Score 53.2; DB 2: Length 4935;  
 Best Local Similarity 44.5%; Pctid No. 0.0016;  
 Matches 211; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

OY 408 ttctctgacctcctaagaagaacagagagagatgagaaataagatacagatgta 467  
 Db 1700 ttactcttcacgttcgctcccaatcccaatcgaacaaagaaagaaacccccc 1759  
 OY 468 tgaagacagagaagagagagatgagaaagagagagagagagagagagagag 527  
 Db 1760 tgcacgtctcctcgtccagagcctatcttctcgaacaaagagagagagagag 1819  
 OY 538 tcaatcagagagagagagagagagagagagagagagagagagagagagag 587  
 Db 1820 gtaacgtcgaacaaagagagagagagagagagagagagagagagagagag 1879  
 OY 588 gaacatcagagagagagagagagagagagagagagagagagagagagagag 647  
 Db 1880 gctacacagagagagagagagagagagagagagagagagagagagagagag 1939  
 OY 648 ccgagagagagagagagagagagagagagagagagagagagagagagag 707  
 Db 1940 agcagagagagagagagagagagagagagagagagagagagagagagag 1999  
 OY 708 aactctcagagagagagagagagagagagagagagagagagagagagagag 767  
 Db 2000 ctctctcagagagagagagagagagagagagagagagagagagagagagag 2059  
 OY 768 atagagagagagagagagagagagagagagagagagagagagagagagag 827  
 Db 2060 atcgagagagagagagagagagagagagagagagagagagagagagagag 2119  
 OY 828 cagagagagagagagagagagagagagagagagagagagagagagagag 881  
 Db 2120 atttgagagagagagagagagagagagagagagagagagagagagagag 2173

## RESULT 9

US-08-810-712-9  
 : Sequence 5, Application US/088107126  
 : Patent No. 6160106  
 : GENERAL INFORMATION:  
 : APPLICANT: Yeda Research and Development Co. LTD  
 : TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
 : TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
 : TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
 : CURRENT FILING DATE: 1997-03-03  
 : EARLIER FILING DATE: 1994-10-12  
 : NUMBER OF SEQ ID NOS: 31  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 9  
 : LENGTH: 5886  
 : TOPOLOGY: Linear  
 : NO. OF FEATURES: 1  
 : ORGANISM: homo sapiens  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (337)..(4605)  
 : US-08-810-712-9

Query Match 11%; Score 53.2; DB 4: Length 5886;  
 Best Local Similarity 44.5%; Pctid No. 0.0016;  
 Matches 211; Conservative 0; Mismatches 263; Indels 0; Gaps 0;  
 OY 408 ttctctgacctcctaagaagaacagagagagatgagaaataagatacagatgta 467

```

Db 1700 ttactcttgacgcttcgctcaaatcccatccgagcaagaagaagaaaccccc 1759
Oy 468 tggagcagcaagaagaaggagatagaataaaggctcccaactcttgctaaagaagg 527
Db 1760 tgcacgcgctcttgacagcgctatctctcgtgcgcaaaagcccttggagacggcct 1819
Oy 528 tcaatccgcgaacatagatctagaagcgagctcgtccctcatgtgtgagctctaaag 587
Db 1820 gtaacgtgacatcaagaaccgagaaagagacgcccctcgtacagcctctgcgcga 1879
Oy 588 gaaacttgtagtcttgaaatgcaactccctctatgaaatgtaattgaacaaagtga 647
Db 1880 gctacacgacatctgtagtgatctctgcgcaaaatgagacgacacttaagtctgagaca 1939
Oy 648 ccgcagaggaagaatgctcctcaccggtcgaagatctggcgaactggtctcaaaa 707
Db 1940 agagcgcgacacttgcctctcatctcgtgaagcgcgtcgtcaatgagatcaaga 1999
Oy 708 taactctacagatacaatctcccatcgaaatgtaagctctcgaggaagaactgcacac 767
Db 2000 cactctcagcaagagtggtcttgctgattatacagacgagcaactacatccctcc 2059
Oy 768 atgctcgcagctatgagcagacgtcctctctcgtacatacagctcgtcgcacaaatggagcc 827
Db 2060 atggtgcatgtgaagaatgycacaatgctctcgtggtgcccctctgaaacaaactgca 2119
Oy 828 cagtgaaatgcgaagaatgatagtggtgcgagacacactgtctgactctccaga 881
Db 2120 attgagacatcccaaaagatgtggcgaagcctcgtacactggtgcgcgcaaaa 2173

```

```

RESULT 10
US-08-181-629A-2
; Sequence 2, Application US/08181629A
; GENERAL INFORMATION:
; PATENT NO. 5472872
; APPLICANT: Genentech, Inc., San Francisco, California
; APPLICANT: Merck & Co., Inc., Kenilworth, New Jersey
; APPLICANT: Merck & Co., Inc., Kenilworth, New Jersey
; TITLE OF INVENTION: Recombinant Cx41 Restriction Endonuclease
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago, Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,629A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Gough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOC. NO.: 3,1504
; TITLE OF INVENTION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear

```

MOTIF TYPE: DNA (genomic)  
US-08-181-629A-2

```

Query Match 1.18; Score 53; DB 1; Length 5496;
Best Local Similarity 43.8%; Pred. No. 0.0017;
Matches 230; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

Oy 509 attctctctaaaaggagatcaatccgagcaagctatagaagaagaacatcgccctt 568
Db 3122 ATTCTCTTAACAGCAGGCTGACCAACCTTATGATACATAAATTTGGCATGATACCCOTTA 3381
Oy 569 cagtggtgagcctcaaaaggaaatctctgagtgcttgaatgcaactccatcaatacagagctt 628
Db 3182 CATTAAGCGGCTCTTTAATGATATGATATGATATGATATGATATGATATGATATGATATGAT 3441
Oy 629 gataatcaacagatgacacgcgcgaagaagaatctctcctcagctcgtcgaagaagtatggg 688
Db 3442 AAGTGTGACCTTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 3501
Oy 689 gatgcacgtgctcacaacaaactctcaagatcaatcgaatcccaatgaaacatctgagacctg 748
Db 3502 CATGAGCATGTGCTGATGACGCGCTTATATGCGGCTCAAAATCTTACATGACACGATAT 3561
Oy 749 caaggagacatgcacatcttaataatgacatgacatgacatgacatgacatgacatgacatg 808
Db 3562 TCGGAGATGTACGACCTTATGATATGATATGATATGATATGATATGATATGATATGATATGAT 3621
Oy 809 ctctgcgacacagtggtcctcgtgaaatgacaaagtctgagtggtgcgacacacatgctgt 868
Db 3622 CTCTGTGACACAGGTGACCTTGTGACGATATGATATGATATGATATGATATGATATGATATGAT 3681
Oy 869 ctgtgtacccagatgtgtgagcacaatatgcaactcgtcgtatgataaagaagaagagat 928
Db 3682 TACCGCGCTTTAATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 3741
Oy 929 attaattctcagagcaacaacaaagagctcctcactgctcagatgagtgagtgatgtctgc 988
Db 3742 ATTGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATAT 1033
Oy 989 aaagatgcagtagaagcttcaatcaaaaagagcgctgaagagac 1033
Db 3802 GATGTGTGATTAATAACATCATGACAGCGATGATATGATATGATATGATATGATATGATATGAT 3846

RESULT 11
US-08-257-073-4
; Sequence 4, Application US/08257073
; GENERAL INFORMATION:
; PATENT NO. 5766597
; APPLICANT: Proctect, Enzo
; APPLICANT: Proctect, Enzo
; APPLICANT: Proctect, Enzo
; TITLE OF INVENTION: VALERIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 510 Fifth Avenue, 25th Floor
; CITY: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783

```

FILING DATE: 11-JUN-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/952,305  
 FILING DATE: 18-MAR-1992  
 PUBLICATION DATA:  
 APPLICATION NUMBER: US 07/672,183  
 FILING DATE: 20-MAR-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Frommer, William S.  
 REGISTRATION NUMBER: 25,506  
 REFERENCE/DOCKET NUMBER: 454310-2570  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 940-3333  
 TELEFAX: (212) 940-3312  
 TELETYPE: 425066 CHTMS  
 INFORMATION FOR SEQ ID NO: 4:  
 LENGTH: 2223 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-257-073-4

Query Match 1.1% Score 52; DB 1; Length 2223;  
 Best Local Similarity 43.5%; Pred. No. 0.0019;  
 Matches 282; Conservative 0; Mismatches 365; Indels 1; Gaps 1;

2789 aaatcacatcaccgaacttaagaagcgcctgcctgcgaacttaataaaatgctgaa 2848  
 1573 aaaaaaaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1632  
 2849 gacacagaagaalatalactctcgtctgaacacatgattgaagaagacacagtc 2908  
 1633 aattatttttaaacacattatattatgcatgaatgatgatgacgttaaa 1692  
 2909 catcagatgctgcctggaacacccagagaagattaaactgccttgagtaacatcg 2968  
 1693 aaattacgtgctcattttaaataattgacacacttaaaatgcttgaacac 1752  
 2969 gataaaccaatgagatagatgctgagagaagaagctgagagcttaatacaga 3028  
 1753 atttaattgatattatgcaaaatatttcaaaaacattgacatgacgttaaa 1812  
 3029 ctctggaataaagaatgagacgaatataaaagaataacctgcgagacactcaga 3088  
 1813 attcaattttttaaacaacatttcaaacgtatttattacattgaattatcaaat 1872  
 3089 caagtaaaagctgagctacatgcgcctgaagaagcctgaagaagatgaagtgcc 3148  
 1873 gcaaaacaaaattatcaaaaatttcaaaaatttcaaaaatttctgaaaaatgct 1932  
 3149 aagagacgaagaagctccagacacacgcgcgaatactcgtcgaagctaaagaaag 3208  
 1933 caaaacacacgaacgaatgcaaaaacacgttgaaacgaattacaaacgaataga 1992  
 3209 gagagagagatgctgaccctgcattgagagagcttgacacccagagagagacga 3268  
 1993 gaaatgcatgcccctccacaaaatttgaagaaacgaacgaacgaacgaacga 2052  
 3269 atacagagatgacatcagctaaatatactcgcacacacgcctcgaagctgaaga 3328  
 2053 aagcaaaacacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacga 2112  
 3329 aacttaaaagcctcgtgagaagaacatgaaagagctacacccagacagacagaag 3387  
 2113 caaaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacga 2172  
 3388 taatacagctgagaagaagcgaagaagctcgaagaagaagacaga 3435  
 2173 gaaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacga 2220

RESULT 12  
 US-08-182-175A-104  
 Sequence 104; Annotation US/08182175A  
 Patent No. 5559223  
 GENERAL INFORMATION:  
 APPLICANT: Saverio Carl Falco  
 APPLICANT: Janet A. Kieker  
 APPLICANT: Janet A. Kieker  
 TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Contain  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: E. I. du Pont de Nemours and Company  
 STREET: 1007 Market Street  
 CITY: Wilmington  
 STATE: Delaware  
 COUNTRY: USA  
 ZIP: 19898  
 COMPUTER READABLE FORM:  
 COMPUTER FILE: floppy disk  
 OPERATING SYSTEM: Microsoft System, 6.0  
 SOFTWARE: Microsoft Word, 4.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/182/175A  
 FILING DATE:  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA: 07/743,006  
 APPLICATION NUMBER: 9 August 1991  
 FILING DATE: 9 August 1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Linda Axemethy Floyd  
 REGISTRATION NUMBER: 33,692  
 REFERENCE/DOCKET NUMBER: BB-1031  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (302) 992-4925  
 TELEFAX: 834720 892-7949  
 INFORMATION FOR SEQ ID NO: 104:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 340 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MODIFIED BASES: DNA (genomic)  
 ORIGINAL SOURCE:  
 STRAIN: E. coli  
 CELL TYPE: DH5 alpha  
 IMMEDIATE SOURCE:  
 CLONE: segment 534 [see 534]  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..340  
 OTHER INFORMATION: /function= "synthetic seed storage protein"  
 OTHER INFORMATION: /product= "protein"  
 OTHER INFORMATION: /gene= "sap"  
 OTHER INFORMATION: /standard\_name= "SAP-534"  
 US-08-182-175A-104

Query Match 1.1% Score 51.6; DB 1; Length 340;  
 Best Local Similarity 48.8%; Pred. No. 0.00092;  
 Matches 141; Conservative 0; Mismatches 149; Indels 0; Gaps 0;  
 2897 tgaagaagaagctggaagatlaaataaagatctggaagaataaagaatgagaaga 3056  
 4 tggacgaacagatgaaagctcgaacgaacgaacgaacgaacgaacgaacgaacga 63  
 3057 tattaagaagaatctggaagacacacacgaacgaacgaacgaacgaacgaacga 3116  
 64 aactcgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacga 123  
 3117 gagagcctgaaagaagaatgagtgagcctcgaagaagaagctgaagaagaagctcga 3176

Db 124 MAACCTGAGAGAAAGATGAGAGCTGTGAGAGCAAGATGAATGCTTGTGAGAAAGA 183  
Oy 3177 ggcgtgaatactgcctaaagtaacaaaaaacccagagagatctgcacctgcagagag 3236  
Db 184 TGAAGAGCTCGAAGAGAGATGAGAGCTGTGAGAGCAAGATGAATGAGAGCTCGAAGAA 243  
Oy 3237 agatgcagccagagagagagagacccagacagatcacgaatgcacacag 3286  
Db 244 AGATGAGAGCTCGAAGAGAGATGAGAGCTGTGAGAGCAAGATGAAG 293

RESULT 13  
PCF-US92-06412-104  
Sequence 104, Application PC/TUS9206412  
GENERAL INFORMATION:  
APPLICANT: Saverio Carl Falco  
APPLICANT: Sharon J. Keeler  
APPLICANT: Janet A. Rice  
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F  
NUMBER OF SEQUENCES: 113  
COMMUNICATED ADDRESS:  
ADDRESSER: E. J. du Pont de Nemours and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: USA  
ZIP: 19898

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCF/US92/06412  
FILING DATE: 19920807  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/743,006  
FILING DATE: 19910819  
ATTORNEY/AGENT INFORMATION:  
NAME: Linda Axamethy Floyd  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: BB-1031  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302) 992-4929  
TELEFAX: (302) 892-7949  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: 1  
STRAIN: E. coli  
IMMEDIATE SOURCE: alpha  
CLONE: segment 534 [seg 534]  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..326  
OTHER INFORMATION: /function="synthetic seed storage protein"  
OTHER INFORMATION: /product="protein"  
OTHER INFORMATION: /gene="sap"  
OTHER INFORMATION: /standard\_name="SSP-534"  
PCR-US92-06412-104

Query Match 1.1%; Score 51.6; DB 5; Length 340;  
Best Local Similarity 48.6%; Pred. No. 0.00092;  
Matches 141; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Oy 2997 tgaagagagactgtgaagataaatacagaagattctggaagaaatgaagatgaagagaa 3056  
Db 4 TGGAGAGAGAGTGTGAGAGAGCTGTGAGAGCAAGATGAATGAGAGCTGTGAGAGAG 63  
Oy 3057 tatataaagaaatctggaagagacacagacacagacagacagacagacagacagacag 3116  
Db 64 AACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123  
Oy 3117 gagagctgaagaaagaaatgaagctgaagacagagagagagagagagagagagagag 3176  
Db 124 MAACCTGAGAGAAAGATGAGAGCTGTGAGAGCAAGATGAATGAGAGCTGTGAGAGAA 183  
Oy 3177 ggcgtgaatactgcctaaagtaacaaaaaacccagagagatctgcacctgcagagag 3236  
Db 184 TGAAGAGCTCGAAGAGAGATGAGAGCTGTGAGAGCAAGATGAATGAGAGCTCGAAGAA 243  
Oy 3237 agatgcagccagagagagagagacccagacagatcacgaatgcacacag 3286  
Db 244 AGATGAGAGCTCGAAGAGAGATGAGAGCTGTGAGAGCAAGATGAAG 293

RESULT 14  
US-08-728-323A-1  
Sequence 1, Application US/08728323A  
Patent No. 5648676  
GENERAL INFORMATION:  
APPLICANT: Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
NUMBER OF SEQUENCES: 21  
COMMUNICATED ADDRESS:  
ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
OPERATING SYSTEM: IBM compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: William J. Russo  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1

Query Match 1.1%; Score 51.6; DB 2; Length 3489;  
Best Local Similarity 45.9%; Pred. No. 0.003;







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 10, 2002, 14:52:41 : Search time 2031.44 Seconds  
(without alignments) 11 updates/sec  
25020.445 Million cell

Title: US-09-750-590-1

Sequence: 1 caggtgtgagcgagcgagat.....atcaaaaaaaaaaaaaaa 4730

Scoring table: IDENTITY\_NGC  
Gapex 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372689281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : 1: em\_estfun:\*  
2: em\_estlin:\*  
3: em\_estlin:\*  
4: em\_estlin:\*  
5: em\_estlin:\*  
6: em\_estlin:\*  
7: em\_estlin:\*  
8: em\_estlin:\*  
9: em\_estlin:\*  
10: em\_estlin:\*  
11: em\_estlin:\*  
12: em\_estlin:\*  
13: em\_estlin:\*  
14: em\_estlin:\*  
15: em\_estlin:\*  
16: em\_estlin:\*  
17: em\_estlin:\*  
18: em\_estlin:\*  
19: em\_estlin:\*  
20: em\_estlin:\*  
21: em\_estlin:\*

pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the best result printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description |
|------------|-------|-------------|--------|----------|-------------|
| 1          | 852   | 18.0        | 135    | AK012454 | Mus muscu   |
| 2          | 611   | 12.6        | 803    | BC120762 | BC120762    |
| 3          | 600   | 12.7        | 848    | BC120762 | BC120762    |
| 4          | 581   | 12.0        | 734    | BC120762 | BC120762    |
| 5          | 562   | 12.0        | 1060   | BC120762 | BC120762    |
| 6          | 554   | 11.7        | 742    | BC120762 | BC120762    |
| 7          | 542   | 11.5        | 929    | BC120762 | BC120762    |
| 8          | 528   | 11.2        | 889    | BC120762 | BC120762    |
| 9          | 528   | 11.0        | 974    | BC120762 | BC120762    |
| 10         | 518   | 11.0        | 666    | BC120762 | BC120762    |

|    |     |      |      |    |          |          |
|----|-----|------|------|----|----------|----------|
| 13 | 515 | 10.9 | 714  | 11 | BC253615 | BC253615 |
| 14 | 513 | 10.9 | 730  | 11 | BC253615 | BC253615 |
| 15 | 513 | 10.9 | 883  | 11 | BC120762 | BC120762 |
| 16 | 510 | 10.8 | 941  | 11 | BC036163 | BC036163 |
| 17 | 506 | 10.7 | 799  | 10 | AA985309 | AA985309 |
| 18 | 504 | 10.7 | 705  | 10 | AA985309 | AA985309 |
| 19 | 490 | 10.4 | 1115 | 11 | AA985309 | AA985309 |
| 20 | 485 | 10.4 | 1115 | 11 | AA985309 | AA985309 |
| 21 | 485 | 10.4 | 1115 | 11 | AA985309 | AA985309 |
| 22 | 453 | 9.6  | 616  | 11 | BC028742 | BC028742 |
| 23 | 437 | 9.3  | 730  | 11 | BC120762 | BC120762 |
| 24 | 429 | 9.1  | 562  | 10 | AA024656 | AA024656 |
| 25 | 428 | 9.1  | 782  | 11 | BC111534 | BC111534 |
| 26 | 427 | 9.0  | 863  | 11 | BC111534 | BC111534 |
| 27 | 424 | 9.0  | 838  | 11 | BC111534 | BC111534 |
| 28 | 412 | 8.7  | 504  | 10 | AA024656 | AA024656 |
| 29 | 393 | 8.4  | 547  | 11 | BC111534 | BC111534 |
| 30 | 387 | 8.2  | 534  | 10 | AA024656 | AA024656 |
| 31 | 386 | 8.2  | 534  | 10 | AA024656 | AA024656 |
| 32 | 378 | 8.0  | 536  | 11 | BC111534 | BC111534 |
| 33 | 378 | 8.0  | 925  | 11 | BC111534 | BC111534 |
| 34 | 378 | 8.0  | 925  | 11 | BC111534 | BC111534 |
| 35 | 377 | 8.0  | 754  | 10 | AA024656 | AA024656 |
| 36 | 372 | 7.9  | 483  | 10 | AA024656 | AA024656 |
| 37 | 370 | 7.9  | 463  | 10 | AA024656 | AA024656 |
| 38 | 370 | 7.8  | 463  | 10 | AA024656 | AA024656 |
| 39 | 367 | 7.8  | 553  | 10 | AA024656 | AA024656 |
| 40 | 367 | 7.8  | 932  | 11 | BC111534 | BC111534 |
| 41 | 367 | 7.8  | 516  | 11 | BC111534 | BC111534 |
| 42 | 365 | 7.7  | 468  | 10 | BC111534 | BC111534 |
| 43 | 364 | 7.7  | 1002 | 11 | BC111534 | BC111534 |
| 44 | 363 | 7.7  | 652  | 11 | BC111534 | BC111534 |
| 45 | 363 | 7.7  | 604  | 11 | BC111534 | BC111534 |

ALIGNMENTS

|            |                                                                                                                                                                                                                                                                                                                                                                         |              |     |             |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|-----|-------------|
| RESULT 1   | AK012454                                                                                                                                                                                                                                                                                                                                                                | 1545 bp mRNA | HNC | 05-JUN-2001 |
| LOCUS      | Mus musculus 11 days embryo cDNA, RIKEN full-length enriched                                                                                                                                                                                                                                                                                                            |              |     |             |
| DEFINITION | AK012454.1 clone:2700059002, full insert sequence.                                                                                                                                                                                                                                                                                                                      |              |     |             |
| ACCESSION  | AK012454.1 GI:12849306                                                                                                                                                                                                                                                                                                                                                  |              |     |             |
| VERSION    | AK012454.1                                                                                                                                                                                                                                                                                                                                                              |              |     |             |
| KEYWORDS   | CAP trapper.                                                                                                                                                                                                                                                                                                                                                            |              |     |             |
| SOURCE     | Mus musculus (strain:C57Bl/6J) 11 days embryo cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library                                                                                                                                                                                                                                                     |              |     |             |
| ORGANISM   | Mus musculus                                                                                                                                                                                                                                                                                                                                                            |              |     |             |
| REFERENCE  | Eukaryote, Metazoa; Chordata; Carnivora; Vertebrata; Euteleostomi; Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.                                                                                                                                                                                                                                             |              |     |             |
| AUTHORS    | Carroll, P., Kondo, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.                                                                                                                                                                                                                                                                                                 |              |     |             |
| TITLE      | High-efficiency full-length cDNA cloning                                                                                                                                                                                                                                                                                                                                |              |     |             |
| JOURNAL    | Methods in enzymology. 303, 19-44 (1999)                                                                                                                                                                                                                                                                                                                                |              |     |             |
| PUBMED     | 99279253                                                                                                                                                                                                                                                                                                                                                                |              |     |             |
| REFERENCE  | 2 (bases 1 to 1545)                                                                                                                                                                                                                                                                                                                                                     |              |     |             |
| AUTHORS    | Carroll, P., Kondo, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.                                                                                                                                                                                                                                                                                                 |              |     |             |
| TITLE      | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes                                                                                                                                                                                                                                      |              |     |             |
| JOURNAL    | Genome research. 10 (10), 1617-1630 (2000)                                                                                                                                                                                                                                                                                                                              |              |     |             |
| PUBMED     | 11042159                                                                                                                                                                                                                                                                                                                                                                |              |     |             |
| REFERENCE  | 3 (bases 1 to 1545)                                                                                                                                                                                                                                                                                                                                                     |              |     |             |
| AUTHORS    | Shibata, K., Itoh, M., Aizawa, K., Nagata, S., Sasaki, N., Carinci, P., Saito, K., Kikuchi, M., Nishi, K., Katsunori, T., Tashiro, H., Itoh, M., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujii, S., Inoue, K., Tozawa, Y., Izawa, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., |              |     |             |









|            |                                                                   |                                                                      |                                                                 |      |
|------------|-------------------------------------------------------------------|----------------------------------------------------------------------|-----------------------------------------------------------------|------|
| D          | b                                                                 | 122                                                                  | AAAACAGCAGCTAAACAACGCTTAAACACTGTGCACCAAAATTCACCGAAGTTAA         | 181  |
| Oy         | 3589                                                              |                                                                      | gaagaagaagaagaaacgtctgtagagaaaataatgccaaagaccttctgagtccttgcaagc | 3648 |
| D          | 182                                                               | GATTGTGAAGAGAGAGCTTGCAACAAAAATAATCCCAAACGACGCTCTCTCAATCTTGGACGT      | 241                                                             |      |
| Oy         | 3649                                                              | acaacactctttgtagagagaagcaagtctgcgtggagaaaggattggaccctcgaaaaaaatc     | 3708                                                            |      |
| D          | 242                                                               | GCAAAATCTTTTGGCAAAACACATCTTCATCTCATGTGCAACAGGTTGAGCCTGTGAAAAATC      | 301                                                             |      |
| Oy         | 3708                                                              | ctcttagtgttaaatctcgaagaacatccaagaagaaacctgaaaacctaagaagagagagatcaga  | 3768                                                            |      |
| D          | 302                                                               | TCTTAATATGCAACATATTAATATCTTAAAGCAACAACTGAAGAAGTAGTCAAAAGCTGTACGC     | 361                                                             |      |
| Oy         | 3769                                                              | gaaagaagaagaagaagcgtggccacacactgcgtggaatgcttcgtagagaaatccgaatcccc    | 3828                                                            |      |
| D          | 362                                                               | GAAGACAGCACAGACAGTGAACCAATCTGCATCAATCTGTGGAGATTCAAAGAACCTCTC         | 421                                                             |      |
| Oy         | 3829                                                              | tgtgcctccctgtagcagctlttgagaggtltaagaagaacatctgagaagaagaagcttgaaatc   | 3888                                                            |      |
| D          | 422                                                               | TGTACCCTTCGCAACACATTTGTGCATATTTAAAGAAAGCATTTGGAGAAAGAACTTGGAATCAT    | 481                                                             |      |
| Oy         | 3889                                                              | aagaagcagaactcttagagagaagaagaagaagaagaagaagcaaaccaaacctgaagaagctacca | 3948                                                            |      |
| D          | 482                                                               | AAAACGCCAGCTTGAGAGAAAGSAAAGAAAGAACCAACCAACAGAGAGAGAGCTCCAA           | 541                                                             |      |
| Oy         | 3949                                                              | actccagctctgagagcttcgaataacatacaagcgtcttaaanaatatlgagactctggagagct   | 4008                                                            |      |
| D          | 542                                                               | AGCTTAATCGCAAGGTCACGAAATTCACAAACAGCATTTAAAAAATTTMAACATCTGAGAGCT      | 601                                                             |      |
| Oy         | 4009                                                              | cgtgtgatcttgctgagataataagaacagaagaagaagcatcttgagagcacagatcttcacgact  | 4068                                                            |      |
| D          | 602                                                               | AGTTCACTCTCTCAATATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT     | 651                                                             |      |
| Oy         | 4069                                                              | aagc--gaaanaattggcacaatcgaatacagaagaagatctgagaagatcagtcga--ggaagattt | 4125                                                            |      |
| D          | 662                                                               | AAACTGAAAAATTTGGCAATCTGGAATGAATAAAGTATGAGCAAGTATGTGATGCAAGCTTTG      | 721                                                             |      |
| Oy         | 4126                                                              | gcaagcacaanaagaagaacgtctctgtcaaaagctgagaagaagact                     | 4170                                                            |      |
| D          | 722                                                               | GATTGTCAAAAGAAAGAGAGAAATCTGGAACAAATATTAATTAACAGAGAAAT                | 766                                                             |      |
| RESULT     | 7                                                                 |                                                                      |                                                                 |      |
| LOCUS      | BG30772                                                           | 742 bp                                                               | mRNA                                                            | EST  |
| DEFINITION | 602827765.F1 NR_MGC_87 Homo sapiens cDNA clone IMAGE:4392374 5'   | 24-JAN-2001                                                          |                                                                 |      |
| ACCESSION  | BG30772.1                                                         | GI:12420213                                                          |                                                                 |      |
| VERSION    | BG30772.1                                                         | GI:12420213                                                          |                                                                 |      |
| KEYWORDS   | EST                                                               |                                                                      |                                                                 |      |
| SOURCE     | human.                                                            |                                                                      |                                                                 |      |
| ORGANISM   | Homo sapiens                                                      |                                                                      |                                                                 |      |
| REFERENCE  | Eukaryotic Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |                                                                      |                                                                 |      |
| TITLE      | Mammalia; Buthera; Primates; Catarrhini; Hominoidea; Homo.        |                                                                      |                                                                 |      |
| JOURNAL    | NR_MGC_http://mgi.nci.nih.gov/                                    |                                                                      |                                                                 |      |
| COMMENT    | Unpublished sequence of Health, Mammalian Gene Collection (MGC).  |                                                                      |                                                                 |      |
|            | Contact: Robert Strusberg, Ph.D.                                  |                                                                      |                                                                 |      |
|            | Email: cseabps-fema11.nih.gov                                     |                                                                      |                                                                 |      |
|            | Tissue Procurement: DCDT/DTP                                      |                                                                      |                                                                 |      |
|            | cDNA Library Preparation: Life Technologies, Inc.                 |                                                                      |                                                                 |      |
|            | cDNA library Arrayed by: Incyte Genomics, Inc.                    |                                                                      |                                                                 |      |
|            | Clone distribution: MGC clone distribution information can be     |                                                                      |                                                                 |      |
|            | found through the I.M.A.G.E. Consortium/LNLN at:                  |                                                                      |                                                                 |      |
|            | http://lmln01.nsl.gov                                             |                                                                      |                                                                 |      |
|            | Plate: LMLN01                                                     |                                                                      |                                                                 |      |
|            | High quality sequence stop 758.                                   |                                                                      |                                                                 |      |
| FEATURES   | Location/Qualifiers                                               |                                                                      |                                                                 |      |
| source     | 1..742                                                            |                                                                      |                                                                 |      |
|            | /organism="Homo sapiens"                                          |                                                                      |                                                                 |      |





[illegible]



Oy 4048 gggagcagagatccgacttaacgcgaataatggccaatctgaatgaagatgaaga 4107  
 Db 361 GAGAGCAAGATTTCTACGCTTAAATGAAAATGGCAATCTGATATGAAATATGAGGA 420  
 Oy 4108 agatagagagaggttttgatggcaaaaagagagactgtccttaagatgaaga 4167  
 Db 421 AGATGTGAGGAACTTTGATGCTCCAAAAAGAGAAATATCTGCAAAAGATGAGAA 480  
 Oy 4168 attgtccttttcgcgaatgaagcaagaatctcaaaagtcagcgagaaagatgagcaatc 4227  
 Db 481 ATTACTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 540  
 Oy 4228 cttcaacacactcagaggttcagagagagagagagagagagagagagagagagagag 4287  
 Db 541 CTTAAACACATCAGAGATTGACAAAGCAATATGCAATATGCTG-TAAACAAATAGAGC 589  
 Oy 4288 aaaaagataaataagataactgaactgaactgaactgaactgaactgaactgaactga 4345  
 Db 600 AAAAAATATATATAGAT-ACCTGACATGCTCTTATGATGTTGCGAAACATTAAACAGGCACTC 658  
 Oy 4346 aatgactcttcgc 4358  
 Db 659 AATGACCTTTCC 671

RESULT 12  
 BGC253615 666 bp mRNA EST 14-MAR-2001  
 DEFINITION 602508850F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4619460 5'  
 LOCUS BGC253615  
 ACCESSION BGC253615  
 VERSION BGC253615.1 GI:13342569  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Tissue: testis; Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 666)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strauberg, Ph.D.  
 Email: c9apbs-remail.nih.gov  
 Tissue Procurement: CLOUTEC Laboratories, Inc.  
 Preparation: CLOUTEC Laboratories, Inc.  
 DNA sequencing by: Invitrogen Corporation  
 CDNA Library: A  
 Clone distribution: MGC clone distribution  
 Found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM1379 row: c column: 13  
 High quality sequence stop: 653.  
 Location/Qualifiers  
 1 666  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4619460"  
 /clone\_1db="NIH\_MGC\_79"  
 /1db\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: placenta; Vector: pDMR-Lib (Clontech);  
 size: 5111 (99cgcctgcgc); Site:2; SfiI (ggcattatggc  
 adaptatg) and 3' adaptors were used in cloning as follows: 5'  
 sequence: 5'-ATTCTACAGCCGACCATTTATGCGC-3' and 3' adaptor  
 (where B = A, C, or G and N = C, G, A, or T). Average  
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
 library."

BASE COUNT  
 ORIGIN

254 a 106 c 150 g 156 t

Query Match 11 08; Score 518.8; DB 11; Length 666;  
 Best Local Similarity 90.18; p=1e-92;  
 Matches 602; Conservative 0; Mismatches 57; Indels 9; Gaps 4;

Oy 1090 ggaacttcaactctacgaagatgacatgaagaatctcaaaagcgagagagactttg 1149  
 Db 1 GAGAGATTTTACGCTTTGAGAGATGATGCAATTTGCAATTTGCAATTTGCAATTTG 59  
 Oy 1150 gaagaaagagagagagagagagagagagagagagagagagagagagagagag 1209  
 Db 60 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119  
 Oy 1210 gaagtcgaactcaagagagagagagagagagagagagagagagagagagagag 1269  
 Db 120 GAAGTCATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 179  
 Oy 1270 gaagagagagagagagagagagagagagagagagagagagagagagagagag 1329  
 Db 180 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 238  
 Oy 1330 ttccagctacagcctgaatgaagaagaatgagagagagagagagagagagagagag 1389  
 Db 239 TTACAGTTACACTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298  
 Oy 1390 aaagctgaagctcccttttgacagcgaagaagaagagagagagagagagagagag 1449  
 Db 299 AAAGCTGACATGCTTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358  
 Oy 1450 gaaagctcgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1509  
 Db 359 TGAGGCTCTGAAAAATGATTAAATTTTGAAGAGATGATTTAGATGATGAGAGATCA 418  
 Oy 1510 ttcca-----ggaagaagatgacttcttcaaaagctcaaatgtacatgacagagc 1563  
 Db 419 TTTCATGACCGAAGAAAGAAATATGCTTTTAAACAGTCAGATGATGAGCAGACTC 478  
 Oy 1564 acaatctactcagagagagagagagagagagagagagagagagagagagagagag 1623  
 Db 479 ACAATGCTCTCCACGATGACCAATGCAATGCAATGCAATGCAATGCAATGCAATG 538  
 Oy 1624 agagctgaactctcaactcaagcactcttcggaagaagcgaatctcaagaagaagat 1683  
 Db 539 GAGAGATTTTACGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598  
 Oy 1684 agaagaagatgaagactctctctctctctctctctctctctctctctctctctct 1743  
 Db 599 AGAAGCAATGCAATCTTGTGTGATGAGCAAAACAGAGCA-TGAGCTCAAAATGA 657  
 Oy 1744 actgagctc 1751  
 Db 658 ACTGAGC 665

RESULT 13  
 BGC253615 714 bp mRNA EST 13-FEB-2001  
 DEFINITION 60236436F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4472416 5'  
 LOCUS BGC253615  
 ACCESSION BGC253615.1 GI:12763431  
 VERSION BGC253615.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Tissue: testis; Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 714)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strauberg, Ph.D.  
 Email: c9apbs-remail.nih.gov





Thu Apr 11 07:43:29 2002

us-09-750-590-1.std.rst

Page 13

Job time: 5933 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: April 10, 2002, 17:05:56 ; Search time 44.15 Seconds

Title: US-09-750-590-2

Sequence: 1 MMSCFSCAPKNRQAADWKK...DVQAAALQITIQMRGLVCSA 1401

Scoring table: BLOSUM62

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```

Maximum DB seq length: 0
Maximum DB seq length: 20000000000

```

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

|    |                                         |
|----|-----------------------------------------|
| 1  | S1D2/gcgdata/geneseq/geneseq/AA1960.DAT |
| 2  | S1D2/gcgdata/geneseq/geneseq/AA1961.DAT |
| 3  | S1D2/gcgdata/geneseq/geneseq/AA1962.DAT |
| 4  | S1D2/gcgdata/geneseq/geneseq/AA1963.DAT |
| 5  | S1D2/gcgdata/geneseq/geneseq/AA1964.DAT |
| 6  | S1D2/gcgdata/geneseq/geneseq/AA1965.DAT |
| 7  | S1D2/gcgdata/geneseq/geneseq/AA1966.DAT |
| 8  | S1D2/gcgdata/geneseq/geneseq/AA1967.DAT |
| 9  | S1D2/gcgdata/geneseq/geneseq/AA1968.DAT |
| 10 | S1D2/gcgdata/geneseq/geneseq/AA1969.DAT |
| 11 | S1D2/gcgdata/geneseq/geneseq/AA1970.DAT |
| 12 | S1D2/gcgdata/geneseq/geneseq/AA1971.DAT |
| 13 | S1D2/gcgdata/geneseq/geneseq/AA1972.DAT |
| 14 | S1D2/gcgdata/geneseq/geneseq/AA1973.DAT |
| 15 | S1D2/gcgdata/geneseq/geneseq/AA1974.DAT |
| 16 | S1D2/gcgdata/geneseq/geneseq/AA1975.DAT |
| 17 | S1D2/gcgdata/geneseq/geneseq/AA1976.DAT |
| 18 | S1D2/gcgdata/geneseq/geneseq/AA1977.DAT |
| 19 | S1D2/gcgdata/geneseq/geneseq/AA1978.DAT |
| 20 | S1D2/gcgdata/geneseq/geneseq/AA1979.DAT |
| 21 | S1D2/gcgdata/geneseq/geneseq/AA2000.DAT |
| 22 | S1D2/gcgdata/geneseq/geneseq/AA2001.DAT |

**Pred. NO.** Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

| Query No. | Score | Match | Length | DB  | ID       | Description        |
|-----------|-------|-------|--------|-----|----------|--------------------|
| 1         | 2934  | 4.7   | 57     | 20  | AA085606 | Sequence of c3ys.  |
| 2         | 2919  | 41.5  | 672    | 20  | AA085606 | Sequence of c3ys.  |
| 3         | 2919  | 41.5  | 672    | 20  | AA085606 | Sequence of c3ys.  |
| 4         | 2919  | 41.5  | 672    | 20  | AA085606 | Sequence of c3ys.  |
| 5         | 2201  | 31.3  | 535    | 22  | AA085650 | Human myofibrin    |
| 6         | 629.5 | 8.9   | 2663   | 3.2 | AA085245 | Human protein sequ |
| 7         | 629.5 | 8.9   | 2663   | 3.2 | AA085245 | Human protein sequ |
| 8         | 628   | 8.9   | 2663   | 3.2 | AA085245 | Human protein sequ |
| 9         | 628   | 8.9   | 2663   | 3.2 | AA085245 | Human protein sequ |
| 10        | 585.5 | 8.3   | 1954   | 19  | AA055241 | Nucleolar/endosome |
| 11        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 12        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 13        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 14        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 15        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 16        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 17        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 18        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 19        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 20        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 21        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 22        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 23        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 24        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 25        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 26        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 27        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 28        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 29        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 30        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 31        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 32        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 33        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 34        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 35        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 36        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 37        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 38        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 39        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 40        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 41        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 42        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 43        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 44        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 45        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 46        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 47        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 48        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 49        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 50        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 51        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 52        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 53        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 54        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 55        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 56        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 57        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 58        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 59        | 582.5 | 8.3   | 1972   | 17  | AA050228 | Human polypeptide  |
| 60        | 582.5 | 8.3   | 1972   |     |          |                    |

|    |       |     |      |    |           |                            |
|----|-------|-----|------|----|-----------|----------------------------|
| 12 | 566   | 8.0 | 1988 | 22 | AAAM0399  | Human poly(epsilon)-lysine |
| 13 | 566   | 8.0 | 1988 | 22 | AAAM1000  | Human poly(epsilon)-lysine |
| 14 | 562.5 | 8.0 | 1979 | 21 | AAAB1811  | Plasmodium falciparum      |
| 15 | 561   | 8.0 | 1392 | 20 | AAV06999  | Restin protein seq         |
| 16 | 561   | 8.0 | 2442 | 21 | AAV17575  | Human cytochrome b         |
| 17 | 561   | 8.0 | 3248 | 17 | AAAG9795  | Kinetochores proteins      |
| 18 | 558   | 7.9 | 1879 | 29 | AAAG5750  | Human protein seq          |
| 19 | 554   | 7.9 | 1879 | 29 | AAAG5750  | Human protein seq          |
| 20 | 540   | 7.7 | 2482 | 16 | AAAT2826  | Human mitotin aml          |
| 21 | 539   | 7.5 | 1789 | 26 | AAAG3596  | Human poly(epsilon)-lysine |
| 22 | 519   | 7.4 | 2192 | 16 | AAAB1321  | LexA/lambda fusion         |
| 23 | 519   | 7.4 | 2272 | 18 | AAAB1731  | LexA/lambda fusion         |
| 24 | 516.5 | 7.3 | 1788 | 22 | AAAG0637  | Human NNN protein          |
| 25 | 514   | 7.3 | 2101 | 15 | AAAR2173  | Human NNN protein          |
| 26 | 507.5 | 7.2 | 2115 | 21 | AAAG9937  | Plasmodium falciparum      |
| 27 | 507   | 7.2 | 1980 | 21 | AAAB1894  | Plasmodium falciparum      |
| 28 | 504.5 | 7.2 | 1562 | 21 | AAAG3923  | Arbidopsis thaliana        |
| 29 | 504   | 7.2 | 1703 | 21 | AAAG5614  | Arbidopsis thaliana        |
| 30 | 504   | 7.2 | 1710 | 21 | AAAG5613  | Arbidopsis thaliana        |
| 31 | 504   | 7.2 | 1744 | 21 | AAAG5612  | Arbidopsis thaliana        |
| 32 | 490.3 | 7.1 | 2101 | 21 | AAAG9736  | Human NNN protein          |
| 33 | 488   | 7.0 | 1521 | 21 | AAAG3923  | Arbidopsis thaliana        |
| 34 | 488   | 6.9 | 1321 | 21 | AAAG3923  | Arbidopsis thaliana        |
| 35 | 480   | 6.8 | 1325 | 26 | AAAG9540  | Human protein seq          |
| 36 | 480   | 6.8 | 1325 | 26 | AAAG9540  | Human protein seq          |
| 37 | 479.5 | 6.8 | 286  | 22 | AAAB9454  | Human RABD5                |
| 38 | 479   | 6.8 | 2139 | 22 | AAAG7128  | Human RABD5                |
| 39 | 477   | 6.8 | 1312 | 19 | AAAG7295  | Human RABD5                |
| 40 | 474   | 6.7 | 1312 | 18 | AAAG7295  | Human RABD5                |
| 41 | 469.5 | 6.7 | 1017 | 22 | AAAB02246 | Domestic mite Bti          |
| 42 | 469.5 | 6.7 | 1296 | 21 | AAAB1592  | Human OPR1                 |
| 43 | 463.5 | 6.6 | 1296 | 11 | AAAB07503 | Mozzelle optical-e         |
| 44 | 463.5 | 6.6 | 1254 | 18 | AAAG4575  | Mozzelle optical-e         |
| 45 | 462   | 6.6 | 5373 | 22 | AAAB1603  | Novel bone marrow          |

## ALIGNMENTS

|         |                                                                         |
|---------|-------------------------------------------------------------------------|
| RESULT  | 1                                                                       |
| AA06996 | AA06996 standard; Protein, 659 AA.                                      |
| ID      | AA06996                                                                 |
| AC      | AA06996                                                                 |
| XX      | 02-JUL-1999 (first entry)                                               |
| DE      | Sequence of C3A5.                                                       |
| XX      |                                                                         |
| XX      | Cancer associated antigen; diagnosis: research; treatment; human;       |
| XX      | breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; |
| XX      | prostate cancer.                                                        |
| OS      | Homo sapiens.                                                           |
| PN      | MO9904265-A2.                                                           |
| XX      |                                                                         |
| XX      | 28-JAN-1999.                                                            |
| XX      |                                                                         |
| XX      | 15-JUL-1998; 98NO-US14679.                                              |
| PR      | 22-JUN-1998; 98US-0102322.                                              |
| PR      | 17-JUL-1997; 97US-0696164.                                              |
| PR      | 10-OCT-1997; 97US-0061599.                                              |
| PR      | 10-OCT-1997; 97US-0061765.                                              |
| PR      | 10-OCT-1997; 97US-0948705.                                              |
| PR      | 11-OCT-1997; 97GB-0021697.                                              |
| PA      | (LUDMG) LUDMG INST CANCER RES.                                          |
| PI      | Chen Y, Gout I, Gure A, O'Hare M, Obara Y, Old LJ;                      |
| PI      | Pfeilschuh M, Sehlin U, Scanlan MJ, Stockert E;                         |
| PI      | Tureci O.                                                               |



QY 903 AEYISLRHEEKMSGLRKMSKRYKQNSAEITAKYKSGSEIVTLHEETIAQKRELIOTIOE 962  
 DB 181 aeyislrheekmsglrkmskrykqnsaeitakyrsgseivtlheetiaqkrelidige 240  
 QY 963 CIRUKTAPIISLEBCEKFKATEKELKESQLOOQYTKNTSEBAKCKKQENOKLKEKELIT 1022  
 DB 241 cikuktapislslebceekfkatekelkeslqseqtkqysveevkknqendkikelt 300  
 QY 1023 LQKDLKDNVHINSEYTERALSKRTBELKROLDLOKTYTEAKKREKLVENAKOTSE 1082  
 DB 301 lqkdlkdnvhinsenyteralskrtbelkroldloktyteakkreklveenakotte 360  
 QY 1083 IIAAORTLLOKQVHPLEOVESLTKSLSGTTTELKEELTKQRCYKFOOTYTOIROMLENO 1142  
 DB 361 iiaaortllokvhpoleovesltkslsgtttelkeeltkqrcykefootytoromleno 420  
 QY 1143 KNSSVPLAEHLQVKEAFEREKGIITKASLRKEBESQNKTEEVSLOSSEIONTKQALKLE 1202  
 DB 421 knsvplaelhqvkeaferekgiitkaslrkebesqnkteevslosseiontkqalkle 480  
 QY 1203 TREYVDSKRYKATKSDLETOISDLNKRLANLNKRYEEVEEYLIANKRELSAANDKRELIH 1262  
 DB 481 treyvdsrkrykatksdletoisdlnkrLANlnkryeeveeyliankrelsaandkrelh 540  
 QY 1263 FSTIEDIRDOQERCKNSLTTITTELORRTOSSANOKEAKONKLTTELNDVERIKQNLNGLS 1322  
 DB 541 fstiedirdoqercknsLttitTELORRTOSSANOKEAKONKLTTELNDVERIKQNLNGLS 600  
 QY 1323 QUTYGSQSPKRSQOLIDSLOQYRSLOQOLADAROROEVIATYRTIILSAAGHNDP 1382  
 DB 601 qutygsqspkrSQOLIDSLOQYRSLOQOLADAROROEVIATYRTIILSAAGHNDP 660  
 QY 1383 YQAAALLQIOMRGSLVC 1399  
 DB 661 yqaallqiiomrgslvc 677  
 RESULT 3  
 AAG64560 standard: Protein: 677 AA.  
 AC AAG64560:  
 DT 18-OCT-2001 (first entry)  
 XX Human myosin heavy chain 74.  
 XX Human: myosin heavy chain 74: disease: cancer; HIV; infection;  
 KW human immunodeficiency virus.  
 OS Homo sapiens.  
 PN CN1296972-A.  
 PA 30-MAY-2001.  
 PF 23-NOV-1999: 99CN-0124081.  
 PR 23-NOV-1999: 99CN-0124081.  
 PA (SHAN-) SHANGHAI BIORONG GENE DEV CO LTD.  
 XX Mao Y, Xie Y.  
 DR MPI: 2001-483699/53.  
 DR N-PSDB: AAH75522.  
 XX Polypeptide-human myosin heavy chain 74 and polynucleotide for coding  
 PT said polypeptide.  
 PS Claim 1: Page 26-27 Disclosure: 36pp: Chinese.  
 CC The invention relates to human myosin heavy chain 74 useful for treating

CC diseases, e.g. cancer, human immunodeficiency virus (HIV) infection.  
 XX Sequence 677 AA:  
 SO Query Match 41.5%: Score 2919; DB 22: Length 677;  
 Best Local Similarity 86.1%: Pred. No. 3.3e-128:  
 Matches 583: Conservative 47; Mismatches 45; Indels 2; Gaps 1:  
 QY 725 KRN-VPLKYSKSHOVIVYDNLKLSLSDYTKIYREKELIMENKALISANYSYL 782  
 DB 1 mknvplkyskshovivydnlklsLsdYtkiYREKELIMENKALISANYSYL 60  
 QY 783 ETVYFIRPERHEKEMALNSNTTELKQLOSLKKNKGDDEKITYSLBNDDKIKMYSHOY 842  
 DB 61 etvyfirperhekemalnsntTELKQLOSLKKNKGDDEKITYSLBNDDKIKMYSHOY 120  
 QY 843 VPKYTERETRLASLIDKTRNENLNYKKCCEDINOPKYKONETIKNLENENNOYK 902  
 DB 121 vpkYteretrlasliDktrnENlNyKKCCEDINOPKYKONETIKNLENENNOYK 180  
 QY 903 AEYISLRHEEKMSGLRKMSKRYKQNSAEITAKYKSGSEIVTLHEETIAQKRELIOTIOE 962  
 DB 181 aeyislrheekmsglrkmskrykqnsaeitakyrsgseivtlheetiaqkrelidige 240  
 QY 963 CIRUKTAPIISLEBCEKFKATEKELKESQLOOQYTKNTSEBAKCKKQENOKLKEKELIT 1022  
 DB 241 cikuktapislslebceekfkatekelkeslqseqtkqysveevkknqendkikelt 300  
 QY 1023 LQKDLKDNVHINSEYTERALSKRTBELKROLDLOKTYTEAKKREKLVENAKOTSE 1082  
 DB 301 lqkdlkdnvhinsenyteralskrtBELKROLDLOKTYTEAKKREKLVENAKOTSE 360  
 QY 1083 IIAAORTLLOKQVHPLEOVESLTKSLSGTTTELKEELTKQRCYKFOOTYTOIROMLENO 1142  
 DB 361 iiaaortllokvhpoleovesltkslsgtttelkeeltkqrcykefootytoromleno 420  
 QY 1143 KNSSVPLAEHLQVKEAFEREKGIITKASLRKEBESQNKTEEVSLOSSEIONTKQALKLE 1202  
 DB 421 knsvplaelhqvkeaferekgiitkaslrkebesqnkteevslosseiontkqalkle 480  
 QY 1203 TREYVDSKRYKATKSDLETOISDLNKRLANLNKRYEEVEEYLIANKRELSAANDKRELIH 1262  
 DB 481 treyvdsrkrykatksdletoisdlnkrLANlnkryeeveeyliankrelsaandkrelh 540  
 QY 1263 FSTIEDIRDOQERCKNSLTTITTELORRTOSSANOKEAKONKLTTELNDVERIKQNLNGLS 1322  
 DB 541 fstiedirdoqercknsLttitTELORRTOSSANOKEAKONKLTTELNDVERIKQNLNGLS 600  
 QY 1323 QUTYGSQSPKRSQOLIDSLOQYRSLOQOLADAROROEVIATYRTIILSAAGHNDP 1382  
 DB 601 qutygsqspkrSQOLIDSLOQYRSLOQOLADAROROEVIATYRTIILSAAGHNDP 660  
 QY 1383 YQAAALLQIOMRGSLVC 1399  
 DB 661 yqaallqiiomrgslvc 677  
 RESULT 4  
 AAB92542 standard: Protein: 535 AA.  
 ID AAB92542:  
 DT 26-JUN-2001 (first entry)  
 XX Human protein sequence SEQ ID NO.10714.  
 DE Human: primer: detection; diagnosis; antisense therapy; gene therapy.  
 XX Homo sapiens.  
 PN EP1074617-A2.

[illegible]

|                 |   |   |      |                                                                                   |      |
|-----------------|---|---|------|-----------------------------------------------------------------------------------|------|
| D               | b |   | 241  | Ingfxfkklkkhnellknrlentqnlkeyjlslncknmslsgamtkvqsmnellan                          | 300  |
| O               | y |   | 936  | YKNSOEELVTLHEEIAAQRRELOJIOETIKLWAPITSLSECRFRFAATEKLKELOLSOO                       | 995  |
| D               | b |   | 301  | ylpkqgeveltlhnelkgkkelldtqeclvyvapilvsfeecerrfkalekeildqlseq                      | 360  |
| O               | y |   | 996  | TOKYNFSEEAQKKCKONDKLKLELTVLQKDLDKRWHTENSVETTERALSRRKTEELNOL                       | 1055 |
| D               | b |   | 361  | dkqysveevkyngnkndgkllkeifrlqpdrtkrctvlyleksemeralsfrtkdelntqj                     | 420  |
| O               | y |   | 1056 | LKLQVYTEAKKEXKXKXKXSELMQAOTLQKHPLAEVYSKSKSSTGTEFLKA                               | 1115 |
| D               | b |   | 421  | kllsqeycvokvkvkvlkenabqkselllvqnllqkphrplogvkaalkngltlenlk                        | 480  |
| O               | y |   | 1116 | BELTKORCEKEDQQVTYDNLQMLFNKSNSSVPALHQLQVEAFERFVGILKASL                             | 1170 |
| D               | b |   | 481  | eelismmqcyeekqegvtvkllhqllennqsnasvplaehhlqleakevegylkksal                        | 535  |
| <b>RESULT_5</b> |   |   |      |                                                                                   |      |
| A               | M | I | D    | AAAM39097 standard; Protein; 2663 AA.                                             |      |
| X               | X |   |      |                                                                                   |      |
| A               | C |   |      | AAAM39097:                                                                        |      |
| X               | X |   |      |                                                                                   |      |
| D               | T |   |      | 22-OCT-2001 (first entry)                                                         |      |
| X               | X |   |      |                                                                                   |      |
| X               | E |   |      | Human polypeptide SEQ ID NO 2242.                                                 |      |
| K               | N |   |      |                                                                                   |      |
| K               | N |   |      | Human, neotrophic; immunosuppressant; cytostatic; gene therapy; cancer;           |      |
| K               | N |   |      | peripheral nervous system; neurodegenerative disease; Alzheimer's dis-            |      |
| K               | N |   |      | Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;               |      |
| K               | N |   |      | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotoxicity;                |      |
| K               | N |   |      | leukemia; thrombolytic; drug screening; arthritis; inflammation;                  |      |
| K               | N |   |      |                                                                                   |      |
| X               | S |   |      | Homo sapiens.                                                                     |      |
| O               | X |   |      |                                                                                   |      |
| P               | N |   |      | NC0200153312-AL.                                                                  |      |
| X               | X |   |      |                                                                                   |      |
| P               | D |   |      | 26-JUL-2001.                                                                      |      |
| X               | X |   |      |                                                                                   |      |
| P               | F |   |      | 26-DEC-2000; 2000MO-US34263.                                                      |      |
| X               | X |   |      |                                                                                   |      |
| P               | R |   |      | 21-JAN-2000; 2000US-0488725.                                                      |      |
| P               | R |   |      | 25-APR-2000; 2000US-055317.                                                       |      |
| P               | R |   |      | 09-JUL-2000; 2000US-0594042.                                                      |      |
| P               | R |   |      | 03-AUG-2000; 2000US-0629312.                                                      |      |
| P               | R |   |      | 14-SEP-2000; 2000US-0663193.                                                      |      |
| P               | R |   |      | 11-OCT-2000; 2000US-0693036.                                                      |      |
| P               | R |   |      | 29-NOV-2000; 2000US-0727344.                                                      |      |
| X               | X |   |      |                                                                                   |      |
| P               | A |   |      | (HYSE-) HYSEQ INC.                                                                |      |
| X               | X |   |      |                                                                                   |      |
| P               | I |   |      | Tang Y.T., Liu C., Asundi V., Chen R., Ma Y., Qian X.B., Ren F., Wang D.;         |      |
| P               | I |   |      | Wang J., Wang Z., Wehrman T., Xu C., Xue A.J., Yang Y., Zhang J.                  |      |
| P               | I |   |      | Zhao Q.H., Zhou P., Goodrich R., Drmanac RT.                                      |      |
| D               | R |   |      | WPI: 2001-442253/AT.                                                              |      |
| D               | R |   |      | N-PsDB: AA158253.                                                                 |      |
| X               | X |   |      |                                                                                   |      |
| P               | T |   |      | Novel nucleic acids and polypeptides, useful for treating disorders               |      |
| P               | T |   |      | such as central nervous system injuries -                                         |      |
| X               | X |   |      |                                                                                   |      |
| P               | S |   |      | Example 4: SEQ ID NO 2242; 1007bp; English.                                       |      |
| X               | X |   |      |                                                                                   |      |
| X               | X |   |      | The invention relates to human nucleic acids (AA157798-AA16369) and               |      |
| X               | X |   |      | the encoded polypeptides (AA157798-AA16369). The nucleotide sequences are useful, |      |
| X               | X |   |      | immunosuppressant and cytostatic activity. The polypeptide sequences are useful,  |      |
| X               | X |   |      | in gene therapy. A composition containing a polypeptide or polynucleotide         |      |
| X               | X |   |      | of the invention may be used to treat diseases of the peripheral nervous          |      |
| X               | X |   |      | system, such as peripheral nervous injuries, peripheral neuropathy and            |      |

localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amphotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

Sequence 2663 AA:

```

Query Match      8.9%; Score 629.5; DB 22; Length 2663;
Best Local Similarity 21.9%; Pred. No. 3.1e-21;
Matches 313; Conservative 265; Mismatches 469; Indels 383; Gaps 56;

Oy 232 NDLDTLLTKTA-----SENSNGKEKAKKGRSLQOQNSMLDIDVTKTSKNGRHA 280
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 844 ngeivlnakeaqqkfdssjgalketajygtelqgektrevgr-----lnemeqkqeqln 898

Oy 281 QN--TODLETENEDEKLERLKIQDQDRIILDVNGLOLDLMEEVN--ADLSESE----- 331
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 899 rdsplqveretkltlelqgtleavkltlqgeqddklt-qlesqigletqkshdihvtn 957

Oy 332 -----KEKLSLLAAKQKQNEESLRTLALKSNFRFESSDGLSSGNF----- 374
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 958 mldtqdeqltnaleel-kqhqe---tlntkfti-----seevrnlhneentigetkdeqf 1009

Oy 375 -----RKEDMLLKQOGMYTDSQSTGTCMVHNSGRSMRLP-----ELALPNQASGE 423
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1010 qkmvgtldkqgdlleakntqltadvkone---lleqgrkfislgelkneiqgmlesviae 1065

Oy 424 NELTKEL-----EAKRTQDSAKQDRLKLDNELAAKVAECALALECERKAE- 471
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1066 keqtkdkenlemltengaelrllygdalkqgeqlvqgdkhmkikeqelstetdliaey 1125

Oy 472 -----DSPEQIKLEDAIKDQKMYEERKQKQKQNRFLAKENGLTDAQNTNRHME 526
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1126 eeklkexsqglqekqgqlnvgeemseuqkklne-----len 1162

Oy 527 LKQDLKDKMYKTEGASAEVGLKLRNOKNEMLVEEFREDESKLMEENKRLQKELSGCELE 586
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1163 lkehlkneleltemetertelqaklne---yeeyk-----slkerayvlnelqkstele 1215

Oy 587 RKRGRKRLTEMEG---OLKDSAKLALSTAEKRFENKSLSNLEKAKKLLDVEREYE 643
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1216 rdhltgylrelelqetqkte-elktis-hlhlkqetelideltrgvesektqllnq--- 1269

Oy 644 RSLNERTRLKRELEMLAKLQAQHVKEPHEBDLK-----SRLEQKSGELGRRTLLTS 695
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1270 -----dlxshktlqgeel-pvlhbeeqqlpnvkvsevcqetmmelilegsettl 1317
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 696 RNRQF--LQRETEYQNDLNLTLQO--VNNLTREKNNVPLAKSRKMKSHOVYDNDLAKKL 751
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1318 kdstltlarlemerlrlnekfgeeqeelsltkerdnlk-llkealevthdqkhefrel 1376

Oy 752 SDVYHRYT--EKKLEMEKLLMENNASLKNVSRLETVPFPE-----RHEKEMALKSNITE 805
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1377 akfgeeqkqgeqlmke---kdneltlvsameq-Elqkdsael-ltletmlyjlskltle 1432

Oy 806 LAKQLSELKAKCGEOKETIYSLMSNNDLKTKMSQIYV--VTHPEITKA----- 854
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1433 shdeemsakexkdqglqgevtlqgeedqlkenltvakhletelaeelavlchalclegeet 1492

Oy 855 ---LSSTLDKTNRELVYKKKCEDIN-----QEVYVAKDNEELLR--NLNEN 896
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1493 lnelrvnsakelelsteikqlgelelnkqlkqelqeykeeqqlkqfisevgekvneqlq 1552

Oy 897 TQNVQVAVYSLREHEEKWSGL-----RKSMMKQV-----DN 928
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1553 lkehlrkaakaaqlsieskmltlnrlqgeeqeqlgmktkeemvtrgqaqlqetrdqlken 1612
  
```

```

Oy 929 SAEILAKYKKSQOE-----ELVTHLESLIAQKRELDTIO-BECLIKRY 966
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1613 lkeaiyvkexsegekeqevglkrlatveiqekmeceelhkqgeqtklnlentetnrlq 1672

Oy 969 APTISLEB-----CERNKFAETEKLKROL500----- 995
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1673 llhneleemsvtkerddlrsvaeltkvzerdqknefrelctlrdekeqlkvlnhike 1732

Oy 996 -----TQKVTSEFEKAKCKQKQNDKATKEILLQKDKMYKNSVTEF--- 1040
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1733 hgetldklrglvseknelelsmqkdehnsndlkqgdllkqgeelrlahmlkveqetldk 1792

Oy 1041 -EVALSRKTEELRNQKLD-----LLOKTFARKERKEL-----VEENAKQTSSELIA 1086
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1793 lrglvsekdklsmqgdlelsmnqkllqgelkanehqltlckdvne-qkksveqk 1852

Oy 1087 OTTLQKQVNYLQVY-----SLAKSSLGRTTLKEKTEL---QRCBEQOQVYQJQKWL 1139
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1853 kklqkqglltleleltnlnqgeelhneleemkvsmkerdnrlrrveetklertdqkkel 1912

Oy 1140 ENKQNSVPLAHQVNEAFKEVGIKASLBEKEEESQNTVEYSKQSEIQNTQKALK 1199
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1913 getkardeiqgelktarmikkehketvdrtekise---ktqlsldqldkdskeqlq 1969

Oy 1200 KLTRENVDSIKY---ATKSDLETOISLNE-----KLAN-LMKRY 1237
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1970 k-----kqelqkqelqlv-ykedvmskxklnemeqllkqgepnlylcksmnhfgllkl 2025

Oy 1238 EYVCEEV-LHARK-----KELASND-----EKELLHFSIE 1266
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2026 hneleelrvakerdelrrlkeslmerdqflactrlremrdqgnhqvkykxrlslsqgq 2085

Oy 1267 GEIKQDQRCDSJLTTELQRIQESAKQIEAKDKMTIELNDV-----RLKQALNG 1320
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2086 hlmeslrfecser---lkelktgysendhyelctrlsldelkelefnrlmklklyay 2141

Oy 1321 LSOVLYGS-GSPKRSQSLDISLQGO-----VRSLOQOGLDADNRHOEV 1363
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2142 vtklkeeqheslnktemndfidevexkqellkllqhlqg---dcvdpasel 2188

RESULT
AM02258
ID AM02258 standard; Protein: 1411 AA.
AC
NC AM02258:
CD
DT 09-MAR-1997 (first entry)
DE
XX Nucleolar/endosomal auto-antigen p162.
XX
XX Auto-antibody; p162; Rheumatic disease; antigen; diagnosis:
XX gene therapy.
XX
XX Homo sapiens.
XX
XX DEL1915J514-C1.
XX
XX 12-SEP-1996.
XX
XX 27-APR-1995; 95DB-1015514.
XX
XX 27-APR-1995; 95DB-1015514.
XX
XX (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
XX
XX Renz W, Seelig HP;
XX WPI: 1996-403153/41.
XX
XX N-PSDB: AAT58751.
XX
XX DNA encoding nucleolar-endosomal auto-antigen - useful for exact
  
```

PT diagnosis of rheumatic disease. In gene therapy and for removal of  
 PT specific auto-antibodies  
 XX Claim 1: Fig 2: 15pp: German.  
 XX Transformed cells can be cultured to produce the antigen p162, for use  
 CC in exact (differential) diagnosis of rheumatic disease. Phenolysis, they  
 CC can detect, in immunoassays, western blots, etc. Phenolysis, they  
 CC specific auto-antibodies. The antigen can be used therapeutically,  
 CC in the removal of auto-antibodies from the circulation, or when  
 CC coupled to a cytotoxicin, the elimination of auto-antibody-  
 CC producing lymphocytes.  
 XX Sequence 1411 AA:  
 50  
 Query Match 8.9%: Score 628: DB 17: Length 1411.  
 Best Local Similarity 21.6%: Pred. No. 1: 8e-21:  
 Matches 294: Conservative 269; Mismatches 444; Indels 356; Gaps 52:  
 QY 220 GHDSSV--YARIGDMIDLTLLTKYASEN--SNKGELM-----KKGCSLQD----- 261  
 DB 67 gndbgngagelnlakrdvdlilrgevdqqlqasleekyeelkkelvnygqlqagkap 126  
 QY 262 -----RMLSDMLDPMVMS--NDR-----HONIDOLEENEDLAKRLKIQDQ 304  
 DB 127 dglvdsaseqlgeqlgeeqtenfltkpmdlfeeqkaqlatcladlckeydeesrlc 186  
 QY 305 RLILDVNLQDLQNEVYVADDESE-----KKKLSILAERKHESRL 351  
 DB 187 eaaeqvvtlrleelikeavtldqldellqgrfledvavllkvelvyqglmdmcler 246  
 QY 352 TFLAKSRFPYFSDHLCSC--SHRKEDMLKQCGMYKTDSCITSTQPMVMSNSKL 408  
 DB 247 eseklkdecklkqsgasealldqgre--lakgqeva-----vygeqql 293  
 QY 409 RFL--LALPNOASYSNELTKELFAMHRTCSMDKORIKLOMLAKHVECKALALDC 466  
 DB 294 kaavnelqlkng-tilten-lllkg-----qdyklk----- 322  
 QY 467 ENVEDSEDOIQDLDAKDQNKHVESGKVAQKQTHFLALKEHLTSDAFCNRL-- 523  
 DB 323 ---ekineesvsk---khqldlqgdldsqqlqerl-----asaelshlnlne 366  
 QY 524 -----NEELKDLQDKMYKEGCAAEVGLKNQIKONE---MLVEFFRDEGKLMEE 572  
 DB 367 lsekgaelqklkeelisevckkyhllkafqlqgqreeknqqlqgeellnqhlkell 426  
 QY 573 NKRLQELSMCELENERKGRKLTPEGQLDLSANLALSTPAEKFNKKSILSN--ELN- 629  
 DB 427 erqlgeahngllkqgqlasaelmdkqaydqlqyl-----srleeqllkvevntelq 481  
 QY 630 ---EKAK-----KILDVEERERKLEWRPLAKJELMLAKKALQD---HY 667  
 DB 482 qldklyqhgqeqqlqgstaklreagqdlqegllqldqkqkqnllelllqgslenl 541  
 QY 668 KPEHIOULSKLEONSGELGKRIITELTSKQNTLOKER---EKVCIDKLLLTQVNNLT 723  
 DB 542 lsektedylaklqagegecal-tnqldkqemltlqeqvqlcaklkqgeshkaqgealn 700  
 QY 724 EM--KNVPLVSEBKKSHDIVDQNKLSLSDVTHKTEKTKLME--KLLM----- 660  
 DB 601 gvgqgkahltraeqdvlslctevneelngqneheskvgajdlqkakkellllsaaka 660  
 QY 771 -----ENR-----SLSKWASLETVFIPPEKHKEKMMALNSNTLTKOLSE 812  
 DB 661 gradlgnhltdanldkqgeqlnkllctqldqyaklqdkqgeqldcshllkveykyls 720  
 QY 812 LMKKGCEOKETYSIMSENNDKRTSMQVPRPKTHIEKLTALSSLTNKLTNPE-- 866  
 DB 722 lsektedylaklqagegecal-tnqldkqemltlqeqvqlcaklkqgeshkaqgealn 700  
 QY 867 -----VDVKKK---CEDIQGFVYKIDENELIKRLMLENTQNOVAVRYSLRHEE----- 913

DB 781 vestrldlqkkaealmeislkqkltlqeeqalldqgdllelqegtklneelnnlqvtvte 840  
 QY 914 -----KMSQLRSMKMKVQDQNSASBILAKYKRSOE---IYTLHEEIAOKRS 956  
 DB 841 lqvknekealmeislkqkltlqeeqalldqgdllelqegtklneelnnlqvtvte 840  
 QY 957 L-----DTIOGCIKAKVAVIISLE---ECERKTRATEKELAEQL--SQTQKRYTSEB- 1005  
 DB 901 lqvgmnltlqeeqelk-----glekeeshnglkelnsmgqllqegnlcklqnekeq 956  
 QY 1006 -----AKKCOENDKLKEITLQDQ-----KDKWVA 1033  
 DB 957 lqgnllnlgasseqkklqlelqgelklavqlktelelnkqqltqaqelaeeekylav 1016  
 QY 1034 IENSYE-----TERALSKTEEL--NR---QLKDL 1058  
 DB 1017 lqnylqgeqetflqgsdflqgsellatrdqlkveekslslaeqdlslmngqlgnqnl 1076  
 QY 1059 LQKTRTEAK-----KEKEKLEWENMKQPSILAOTLQKQYH---PLDQVSLKKS 1107  
 DB 1077 lqelkakkellqgsaekqqlqerckelldqldkqsklkeelvncklslaeleek 1136  
 QY 1108 SQTLETLKELKTCORCYKEEOQVTVLQOMLNQNSVPLASHLQVKEAFKE---VW 1164  
 DB 1137 ekeltklmeikshlkslkeellnlkdaqllllqqlklqgkdsll--kaevqekrnq 1194  
 QY 1165 ITRKSLREKREESONK--TEFYSKLOEIQNTQKALKITLEFVQDLSKTKRYTSLOEQ 1223  
 DB 1195 llkqykkeeeelklfeleekellhalelkevegnkheeme-----aklmql 1243  
 QY 1244 SLDLEMLANLNRRYECVPLMAKKELSAKDEKELIHFSIEDKIQDQ---ERC--- 1276  
 DB 1244 talneldlgvkewqssqtrvseleqlddlqgelavleavqmgndferrallerclyge 1303  
 QY 1277 -----DKSTTRTLQRR-----IQSQAQOIFAKMCKRYTELLN 1309  
 DB 1304 gelekldqkvlqelqrlldlctlaeqqlqgengqlqkthqaln 1346  
 RESULT 7  
 ID AA040883 standard; Protein: 2688 AA.  
 XX AA040883:  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SPO ID NO 5814.  
 XX  
 KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW paraneoplastic; paraneoplastic; paraneoplastic; paraneoplastic;  
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;  
 KW chemokine; chemokine; chemokine; chemokine;  
 KW leukemia.  
 XX  
 OS Homo sapiens.  
 PN WC0200153312-A1.  
 XX  
 PR 26-JUL-2001.  
 XX  
 PF 26-DEC-2000: 2000W0-US34263.  
 XX  
 PR 21-JAN-2000: 2000US-0488725.  
 PR 25-APR-2000: 2000US-0553117.  
 PR 09-JUL-2000: 2000US-0598042.  
 PR 13-OCT-2000: 2000US-0620312.  
 PR 14-SEP-2000: 2000US-0623456.  
 PR 14-SEP-2000: 2000US-0623456.  
 PR 19-OCT-2000: 2000US-0693036.  
 PR 29-NOV-2000: 2000US-0727344.







|  |          |                                                                           |                                                                  |      |
|--|----------|---------------------------------------------------------------------------|------------------------------------------------------------------|------|
|  | OY       | 1231                                                                      | ANKRKEE--VCEEVLTAAK--KEELSAADKKELHPFISDIEKHQDGRCKS-----          | 1279 |
|  | Dd       | 1698                                                                      | rneekakkaiddamaameelkkedtsahlernkmoqtkldghldeaqtaikgg            | 1757 |
|  | OY       | 1280                                                                      | LTTTELQRIOESAOJEAONDKRTTELLNDVERLAKNLNLSQLTVSSGSPSK---ROS        | 1336 |
|  | Dd       | 1758                                                                      | kklqqltselemleneledexkrime---svkmrkserrtikeltgyleedkhnlvrlq      | 1814 |
|  | OY       | 1337                                                                      | QLIDSLQOQYSRLQGLDLAPNRQDEVLATYY--THLSAAGQMD                      | 1380 |
|  | Dd       | 1815                                                                      | dlydklglkykayrqaceeqantlnskfxfqhedeerad                          | 1860 |
|  | RESULT_9 |                                                                           |                                                                  |      |
|  | AA01632  |                                                                           |                                                                  |      |
|  | AA01632  |                                                                           | standard; Protein: 2954 AA.                                      |      |
|  | AC       | AA01632:                                                                  |                                                                  |      |
|  | XZ       | 22-JUN-1999                                                               | (first entry)                                                    |      |
|  | Dd       |                                                                           | Amino acid sequence of centromere-associated protein-E (CENP-E). |      |
|  | KW       | CENP-E; centromere-associated protein-E; ATPase activity;                 |                                                                  |      |
|  | KW       | plus end-directed microtubule motor activity; chromosome congression;     |                                                                  |      |
|  | KW       | microtubule binding activity; chromosome movement; mitosis;               |                                                                  |      |
|  | KW       | cell division; cancer; metastasis; vascular malformation;                 |                                                                  |      |
|  | KW       | inflammatory disease; resenosis; angiogenesis; hypertension;              |                                                                  |      |
|  | KW       | resenosis; fungal infection; solphase; tumor; kinase;                     |                                                                  |      |
|  | KW       | insecticide; plant growth regulator; activator; cancer cell marker.       |                                                                  |      |
|  | OS       | Xenopus sp.                                                               |                                                                  |      |
|  | PX       | M09913061-AI.                                                             |                                                                  |      |
|  | PX       | 18-MAR-1999.                                                              |                                                                  |      |
|  | PE       | 10-SEP-1998: 98NO-US19231.                                                |                                                                  |      |
|  | PR       | 11-SEP-1997: 97US-0058645.                                                |                                                                  |      |
|  | PA       | (RECC ) UNIV CALIFORNIA.                                                  |                                                                  |      |
|  | P1       | Cleveland DM, Goldstein LSB, Sakowicz R, Wood KW;                         |                                                                  |      |
|  | P2       | WPI: 1999-220933/19.                                                      |                                                                  |      |
|  | DR       | N-PDB: AAX26819.                                                          |                                                                  |      |
|  | PT       | Centromere-associated protein-E and related nucleic acid                  |                                                                  |      |
|  | PS       | Claim 5; Page 66-67: 77pp; English.                                       |                                                                  |      |
|  | XX       | The present sequence represents CENP-E (centromere-associated protein-E)  |                                                                  |      |
|  | CC       | of Xenopus. The protein has at least one of plus end-directed microtubule |                                                                  |      |
|  | CC       | motor activity, ATPase (adenosine triphosphatase) activity and            |                                                                  |      |
|  | CC       | movement toward                                                           |                                                                  |      |
|  | CC       | of chromosomes during mitosis. These ends are essential for congression   |                                                                  |      |
|  | CC       | cell proliferation. Agents that modulate CENP-E activity thus control     |                                                                  |      |
|  | CC       | theputative, biogicalnural and diagnostic agents, e.g for treatment       |                                                                  |      |
|  | CC       | of unwanted cell proliferation (typical of many examples are tumors and   |                                                                  |      |
|  | CC       | metastases; vascular malfunction); inflammatory and immune diseases;      |                                                                  |      |
|  | CC       | angiogenesis; hypertension; resenosis; and fungal infections), also as    |                                                                  |      |
|  | CC       | plant protection agents (selective herbicides, fungicides and             |                                                                  |      |
|  | CC       | yeasts. CENP-E is also a diagnostic marker for dividing cells, including  |                                                                  |      |
|  | CC       | cancer cells.                                                             |                                                                  |      |
|  | SQ       | Sequence 2954 AA:                                                         |                                                                  |      |

Query Match      8.3%    Score 585.5; DB 20;    Length 2954;  
Best Local Similarity    20.3%;    pred.: NO. 3.9e-19;

| Matches | 334: Conservative                                                  | 302: Mismatches | 565: Indels | 443: Gaps |
|---------|--------------------------------------------------------------------|-----------------|-------------|-----------|
| OY      | 30 ERSDVEKVSIIIAKKGVNGLVDEERAPFVAVMSKINLELNAIIILIGVDTTSPTAQR       | 89              |             |           |
| Db      | 614 EKVGVYKSMVIGAGSNASDGLVSGVGRKJSSHECIE                           | 656             |             |           |
| OY      | 90 NALHLAKYGMALCE-ORLIDYKCPRIYVDCGRRLALMDAPCSGTLIDBKAQV            | 148             |             |           |
| Db      | 657 -----hknleqkivdldeeflenlkkkaendkqcsqsdmeiq-lce---alm           | 700             |             |           |
| OY      | 149 AKDVGRTPLVATLQMKRPTLCOLLIDG-----ADVMSDKQNTMTLMCEVGCRA          | 202             |             |           |
| Db      | 703 deanaalegal-----mrdfnhlimeklrteladlerlekngeet-----ne           | 748             |             |           |
| OY      | 203 VEVLAKKKA-----DPTMLDLGHSYXARICM-DITLTKRAENSKKRE                | 251             |             |           |
| Db      | 749 fellekeqkeheqheqleltheqsklivamey-----npleiedetkxllikeqqlale    | 805             |             |           |
| OY      | 252 LAKKPSLTO-----RNLSONLMDVNTKSNQRFHONODILEYEMKLRKRLTGOEGRYLD     | 309             |             |           |
| Db      | 806 ltrtdnldqkqvtrfd-----svmgdeeklceqlfqlkqslsdaevtrdaakeqls       | 869             |             |           |
| OY      | 310 KVVAGLQIDLVNMYADLDESEKKEIIIAKKKQKHQESLRTITELKSNFTRESNIG        | 362             |             |           |
| Db      | 863 e-----nltekmdeslsmvnyqkeasell-----ekq-----leleksynlmead-----   | 905             |             |           |
| OY      | 370 SCSHRREMLKKGQGVNTDSCSTGSGVNMOSRMRYLEL-----                     | 413             |             |           |
| Db      | 906 -----lqklsqslsefnlgnl-gllaqkypvrdllsvaleklysefqlkeale          | 954             |             |           |
| OY      | 414 -----ALPQNS-YSNELIKLEELKEMKTRFCSQMDKQRLKONELKRLVBEKCALDECEV    | 469             |             |           |
| Db      | 955 ekhalemevtrclseqlfipneveclcknglshaseellmlkqgehsa-----slsqyell  | 1010            |             |           |
| OY      | 470 KESDSDIQILDELDLDVQKMYESEKVCYQKQCHNTFLAKENHTGDAQNTNRIMAEELD     | 529             |             |           |
| Db      | 1011 meqesqdeqlvlethqlq-----ekydrgeeqylemk-nhndlfekyrlmkeasad      | 1065            |             |           |
| OY      | 530 QLKDMK-VYECASAKYKGLRNLQNIQKQKLELDEKQKSKMEK-----RLQNEISMCE      | 584             |             |           |
| Db      | 1063 lllremndlgimevev-----kladethaleldeidqellnekyqlfgamqellftrp    | 1112            |             |           |
| OY      | 585 LEERKQRKLTLMQCO-----IKQSLAKLALSTIPKFEFNKMSLSTENELKMKELLDYER    | 640             |             |           |
| Db      | 1118 ldsdipkskivagsgdplndynhljal-----sternlmvclaternalekveid--     | 1171            |             |           |
| OY      | 641 EYERSLMERPLKLEELKELKAKLADQIVKREERDQLSRLQSKSGELANRTLETSTKNOTL   | 700             |             |           |
| Db      | 1172 -----lntqqlqql-----eqsklsdqlqrqqldegeevllllemllnglchds        | 1211            |             |           |
| OY      | 701 QKREIKCYLDKMLLTGOOVNMLTTEKMKVNVPLKVS-----BEYKSHQVLYVDQKRLSDV   | 754             |             |           |
| Db      | 1218 qlseklqglehlevetkqqlqgleekmkfitemdqqlvfeelheebshlksqslseqls   | 1273            |             |           |
| OY      | 755 THRTTEKLELMEKLTLMENASLSKNSVSLTEVAPREERKEEMK-ALKSNITELKQULSEL   | 817             |             |           |
| Db      | 1278 lqetqdelraaqeell-----rqqlvdeqlfqqqldeqqldeqqlasqp             | 1315            |             |           |
| OY      | 814 NKKSGEQRKITSYMSBNQDKKTKTSYQVUPVKNHEKT-----ALSSSTD-----RTN      | 863             |             |           |
| Db      | 1316 ndhavanagaq-slgvemsilqsemlr-----getdelqscslavaleelltrahvksve  | 1366            |             |           |
| OY      | 864 REIDYKAKKCDINDQPKTKKQNTLAKNLELNTQ-----NOVK-AETISLRSH-----EEK   | 914             |             |           |
| Db      | 1369 genleltklqglekelslgseeseevlkeamlenlkednmlkqeqyasekqlqfdeev    | 1420            |             |           |
| OY      | 915 MSLIKSKMKKQVONSLELLKY-----                                     | 936             |             |           |
| Db      | 1429 ftsqqlqqldeevlqkqqlkaeeeltekdrqylelvgvsnthlveqltrqpedeeds     | 981             |             |           |
| OY      | 937 -----KSSQK-EYVLELAKKQKRLTQLOSTKIKLY-----PISLESEKRF             | 988             |             |           |
| Db      | 1489 ldrtrseemvllqveklterovqlletrqle-eklelsnllslqlkpmvtrllldlqokll | 1547            |             |           |



Db 1426 gqeldlvtaldnqrlqvlanelek-----qtkedqllae-----eknisakya----- 1468  
 Qy 853 TALSSTLRKTRNRELVDKRCEDNOEVRKIKRENEELIKKNLEPMONQVAKYISLREHR 912  
 Db 1469 -----derdraeeareek-etkatslatralealeakheelerltnlkhae----- 1512  
 Qy 913 EKMSGLSRKSKRVQVDSABETLAKKYSOEELVTLEHEITAOKREL-DTIOECIKLRVAP1 971  
 Db 1513 --medvyskxdyghvnhel-----eskratleqneemtkqleesedvgatcedakrlle 1566  
 Qy 972 TSLEF-----EEERKPRATE-----RELKEOLSGOOTOXKNTNPEPKAKGQEMDKLKKLLILMO 1024  
 Db 1567 vnmqalkqgfedrlqardeqneekrrlqrlqrlheyleledeledeqrqlalaanaakkk-----le 1623  
 Qy 1025 KQLDKMNVHLNENYTERALSRTKEEYLNRQNLQKLTQKRTAK-----KEKEK 1071  
 Db 1624 gdlldlelqadelslkgrteaalkqrlrkqgmkdgrfelddaradelafatskenekak 1683  
 Qy 1072 LVEENKAKOTSTILAAQTLQKQVPLEO---VSLKSLSGSTIETLEKELTQKRCLEK 1128  
 Db 1684 slendlmqldelaaeeraqyrdlekeelaeelaaslsqy-ntlqde-----krl----- 1734  
 Qy 1129 OQYTOLOKMLENOKNSVPLAEHIO---VKEAFKEKVGITRKASREKEESQNKREY 1184  
 Db 1735 earlaqlleeeleegmeamedrvrkalklqeqalselateretqkhesaqrqlerq 1794  
 Qy 1185 SKIQSEONTKQALKLETRREVYDLSKYKATKSDLETOISDLNKLANLNKRYEEVCEY 1244  
 Db 1795 kdlstsklqevagvkr-----alkstsvaalalelqleeqvegealeq----- 1836  
 Qy 1245 LNAKKKEELSAKDE---ELAHFSIEDEIKDO-OECDCPSLTITTELQORIOEASQ---1 1297  
 Db 1839 --aatsklskqdkkiklevllqvederkaeqykeaqeagntkvkqlrkqlleeeesqql 1896  
 Qy 1298 EAKDKRTITELLANDV---ERLKQALNGSL-SQLTYTSGS---PSKR 1334  
 Db 1897 nmnrtrklqeldeaesneemqrevalkskltrrgneastvpsrr 1941  
 RESULT 11  
 AARI0534 standard; Protein: 1427 AA.  
 ID AARI0534 standard; Protein: 1427 AA.  
 AC AARI0534:  
 PY 12-APR-1991 (first entry)  
 DE Human 160kD mediator of inflammation protein.  
 KW Mediator of Inflammation; cytokine; Hodgkin's lymphoma; MRP-160.  
 OS Homo sapiens.  
 PN EPR12050-A.  
 PA 06-FEB-1991.  
 PR 05-JUL-1989; 89GB-0015414.  
 PA (CIBA ) CIBA GEIGY AG.  
 PA Odink KG, Tarcsay L, Bruggen J, Wiesendanger W, Cerletti N;  
 PI Sörg C, Demolt-Feetters C, Delibade J;  
 XX WPI: 1991-038913/06.  
 DR N-PSDB: AAO10378.  
 XX 160 kD human polypeptide mediator or precursor of inflammation -  
 PT diagnose chronic inflammation and Hodgkins lymphoma  
 XX

PS Claim 3; Page 32; 47pp: English.  
 XX  
 CC The protein is a cytokine used to treat chronic inflammatory  
 CC conditions. It is prepared by chromatographically purifying an  
 CC optionally pre-purified cell extract, cell supernatant or cell  
 CC filtrate of stimulated normal human leucocytes or human embryonic  
 CC epithelial lung cells. Alternatively, the protein can be produced  
 CC by microorganisms or continuous mammalian cell culture. The  
 CC protein is a polypeptide fragment from amino acids 878-1427 and covers the  
 CC polypeptide fragment from amino acids 878-1427 and derivatives of  
 CC the protein in which the amino and/or hydroxyl functions are  
 CC glycosylated or acetylated and have mol. wt.s of 180 and 140kD,  
 CC respectively.  
 CC  
 SQ Sequence 1427 AA;  
 Query Match 8.2%; Score 580.5; DN 12; Length 1427;  
 Match Local similarity 22.9%; Ref. No. 3-19;  
 Matches 169; Conservative 236; Mismatches 424; Indels 245; Gaps 46;  
 Qy 254 KKGSLQQRNLSQMDLVN-----TSQNRHONTOOLEEHEDELERIRPTO 301  
 Db 308 krpssas--slssmsvsvavcsasprtgllltstsrqyarklsgtatlqaelqkxqqlle 365  
 Qy 302 Q-----EQRLLDKVNGLDQDLNEEYVAD-----LESEKELSLAAKQHE 347  
 Db 366 qlleerdlerevakaslvqelqelqelqldghqqlvlelakmqdlamveadeeky 425  
 Qy 348 ESLRTTELKSRFRKPRFSDHLSGSHPRKEDMLKQOQVHTDSCSTPCVPHVQSRW 407  
 Db 426 ellqqlleeklkvedlq-----frveesalkqdl-----lqtkl-----ehaa 465  
 Qy 408 LRPTELALPNQASYSBNEELKELAEAKRTFCSAKODLKLONELIAHKVACAL--ALP 465  
 Db 466 lveelqslleekkalq--lqreldelavtseskrimelekdalqylveelrrlle 523  
 Qy 466 CERVKEDSDPQIKQLEADLADVDQKRYREDESKQVQKQNHGLKLFHSLTDDAAGNHR1 523  
 Db 524 snkpgdvmaiajlqel-sslqekl--evtrtdqqltelkqhgrcehkyelka 579  
 Qy 524 ---MEELKDQDKMKVYEGASAE-----VGRLLNIOIKONBLVEERRD----- 565  
 Db 580 lylalelksenelskshankensadvaykshletalsbhqmeeklvsfasklg 639  
 Qy 566 -----EKLMEENKRL--OKELSKCELPRE-KRGRITLPMQGLDLSKSLASIPAE 615  
 Db 640 tetaefaelkqlekmrldyqlnealnpqgdseraahakeaa---lrrkl-mkyike 694  
 Qy 616 KFEENKSLSNELNEKARKLIDVERERSLNTRPLKRLLENLAKLQVYKPREH--E 673  
 Db 674 kenslelrlskldkaedhlvemedclnk-lqeeelkykvelvqklskexlaas 753  
 Qy 674 QLKS-----RLDQSEGLKRTITELSKNOTLOKETELKVCIDMKLKTDOOVNNITTE 724  
 Db 754 qlkaleekllldalrlkaseegksmkrlrqleaeqqlkhltelekaesskasente 813  
 Qy 725 MKNVPRLKVESEMKKSHDV--IVODLNKTSLVNTHKTRKLEPKERLMEKNSLSKRVNRL 782  
 Db 814 lqreldelavtseskrimelekdalqylveelrrlle 870  
 Qy 783 ETVTPRENNKEMALKSNITELKQSLSEKQCEQOKRYSLMS-----NNDLKTTM 838  
 Db 871 -----hkeefgmfnlsadeklrlndimeaklfedderseqllylakeklnidieelm 923  
 Qy 839 SHQVPLKTHIELKTLSTLQKTRALVDVKKSCDINDEPVKIKDEMLIKLNLNENQ 898  
 Db 924 -----kmgdnasqglkmmddrlkxrdveelqklklnknaasflqgslehm 971  
 Qy 899 NOVKAETI---SLREHE-----KNSGLRSKRVQVDSABETLAKKYSOEELVTLEHE 948  
 Db 972 -lvkaeqsqdaaklkheekelerkslstelekmsnmgelakaygetasetthe 1030







```

Oy 518 TGNRLMEELDOLDKMVKESASAEVOKIARNOI-----KONEMLVEFRKEDCKIMEN 573
Db 294 --kneinaInndhlekekkylleyeleenqgldllnkqeketekecekeceke 351
Oy 574 KR-----LOKEL-----SKELETERKRGKLFEMGOLNLSAKLA----- 609
Db 352 kekeydtlllxeladekslllekvnslkveymdelekrehnflimedqkldklnsvfknnq 411
Oy 610 LSTPAKCFPNKSLIS-----NELNKKAKLIDVEYREYSLN- 647
Db 412 lvykceelknltkelekekeldkdeanskeeklnlqngnekkqglefrnkheeq 471
Oy 648 -----STRPL-----KRELENNKKAKLAQHV-----KPEREBOIK- 676
Db 472 lkeelkeevyktlcteqeqemvdlqyqelddqyqeqnqaslelselskvekygn 531
Oy 677 -----SRLEKSGELKRTPELRSN-----OTLKEIKYVCDNK 712
Db 532 cyleeelnleketeekeylnlqnyneelnlnndlnmnglnkltmnglsleltkdn 591
Oy 713 LITGOVNNLTTEKKNVPLKYS-----EMKSHOVI-----VDDANKKLSOVTHKATK- 761
Db 592 llnqgldklnmnglnsklnslngvndlnkeekflnnglvdsnqldlltrknekan 651
Oy 762 -----KLEPMKLEKMSLSKV-----SRLEVIPIPERNEK 794
Db 652 kmlegenykgeme--llrgn-lksenllndeovcdlkkslskesemkm-----keedk 707
Oy 795 EWMALSN-----ITELKOLSE--LNMKGCEDEKITSLSMESEKODAKTM 838
Db 708 klaelkdcdvrlremnekedlnmlkeeyeklnltkegednltlkegednltl 767
Oy 839 SHOVY-----PYKTHETKMLSTLDKTRRELVYDKKCDINDOFVKIKDE 886
Db 768 keeyeklnlmkeeyeklnltnegeeklnltnegeeklnltmkn-----eegeekmclneq 825
Oy 887 NE-----ILKNLENTONOKYA-----EYISAREHEKMSGLKSMKKYOD--N 928
Db 828 nedkmnskeeyenqlnglnsmnelklkdvneyl-----eevdklvtldkklkqgdkeln 882
Oy 929 SAEILAKKKSQSEIYVLMEDIAOK--RELDTIQDCIKL--KYAPITSLERCE----- 978
Db 883 yehlkahak--eglllemedkcgdkysdlyekyiklkslcmllmceddlened 940
Oy 979 -----BKFKATEKLEKOLSOQOKYKNSSEBAK-----KOKEDMDIKRETL 1021
Db 941 ltrfleeylnmngklykeveekhnkfnllkskckfrfnsiedkshelkhekdll 1000
Oy 1022 TLONDKCKVHENTSERALSRKTEELNQLDKOKYTAKEKEKLEVEENAKOS 1081
Db 1001 skheleekn-----Kkllelnmdklrqdelvlykk-----gsnaqgd 1040
Oy 1082 EITLAQTLORO-----HPLLEGOVYSKLSGRTIKTEL-KYKOKCEKE 1128
Db 1041 hktksawlilkdskekldkenglnveheekdkkk--odefrlneelvykkllylnk 1099
Oy 1129 QOQTVTQLOMLKNSNSVPLASHLOVNE-----FEKEVGYTASJAPKSEESONAKTE 1182
Db 1100 koplqngldklkldlnstlnegmckvleellloyleeInsrftslqjlnelcsilt 1159
Oy 1183 EVSKLOSSEIONTKALKLETRBVDSLKXKATRSOLOTO-----ISDMLELNANKRDE 1239
Db 1160 eweelnmknleleennkl--nlvdsqk--klkldvekketelknkqlkmgld 1215
Oy 1240 VCEEVYLAKKKELSANDEKELLFESIDREIK--DOOE-----RODSLSYTRTLELROROS 1293
Db 1216 lneev-----eklnneelnellcyndlnmktldmkenmlmkldenedlkkmsklddm 1269
Oy 1294 ANQIDAKMDKITELIDVBERLO 1316
Db 1270 ekelyredekknlnlelnklk 1292

```

```

RESULT: 15
AAV06999
AAV06999 standard; Protein: 1392 AA.
AC AAV06999;
XX
XX 02-JUL-1999 (first entry)
XX
XX Reslin protein sequence.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer.
XX
XX Homo sapiens.
XX
XX NC0904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98MO-0514679.
XX
XX 22-JUN-1998; 98US-0102322.
XX 17-JUL-1997; 97US-0896184.
XX 10-OCT-1997; 97US-0061599.
XX 10-OCT-1997; 97US-0061765.
XX 10-OCT-1997; 97US-0948705.
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Cure A, O'Hare M, Obata Y, Old LJ,
XX Pfrendschuh M, Sahin U, Scanlan MJ, Stockert E,
XX Tureci O;
XX
XX WPI: 1999-132448/11.
XX N-7SDS; AAK4200.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
XX isolated using sera from cancer patients - for the development of
XX for the diagnosis, monitoring or treatment of cancers
XX
XX Example 8: Page 783-786; 787pp: English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample with a probe derived from a subject with an agent that specifically
XX binds to the NAM, an expression product and a detection reagent; an expression
XX product complexed with an HLA molecule; and (b) determining an expression
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
XX
XX Sequence 1392 AA:
XX
XX Query Match 8.0%; Score 561; DB 20; Length 1992;
XX E-Value 2.2e-38; Ident 22.3%; Posit 22.3%; Mismatches 386; Indels 294; Gaps 44;
XX Matches 962; Conserved 231; Mismatches 386; Indels 294; Gaps 44;
XX
XX 238 LKRTASNSNK--GREILKKGTSLOQNSQMDVENRK-----S 275
XX 336 lteasryarkgtalgaekqheqlleedleaeavakshvgelegelala 395
XX 276 NREHONIODLETENEDK-----ERLRKIOQBORILLDKVNLQLOLNEVAVY 324
XX 396 rgnldqnyvleakmgqlrtmveaadrekvellnqleekr-----tvedlqfrveeslt 451

```

Thu Apr 11 07:43:30 2002

us-09-750-590-2.std.rag

Page 16

OY 325 ADDE---SEKELAKS---LAKEKOHESITETALAKSRFYESDHLGSGSHRE 377  
Db 452 kdelvavsekerimalelekdalivgevelrrieseokp-----agdvdmlalilqel 506  
OY 378 DMLKOGCOWTSDQSTGTOMPVHOSRMLRPLRLALPNQASYSENSEILKKELEAMRTF 437  
Db 507 sliqelavtrchrtelitelkenfigar-----eetbqkelkalyla 548  
OY 438 CSMAKODRIKLOVELAKVAKESKALALCEVNEEDSDQI---KOLEDALKDOKRWYES 494  
Db 549 tekiskeneslikekileh-----ankensdylavakletelashqameel 595  
OY 495 EGKVKOMQHFLAKEHLTSDMATGNHRLMELELDQKDKKRYEGASAEVGLRNDQKO 554  
Db 596 k-----vafakigjeltelae-----faelktqlekarldyq---helenlqng--- 634  
OY 555 NPMILVEFRDECKIMEENKRILOKELSMCELEPERKGNKILTMEGOLADSLALASTPA 614  
Db 635 -----qdsoratahakemaa-----lrakl-mkylk 658  
OY 615 EKFENMKSLTSELNEKAKKLIDVREVERSLNTRPLKRELENLKAKLAQHKPEER-- 672  
Db 659 ekemialelreklkdaedqilvemedlntk-lqaeelivkelevlqakeneqclkvldntf 717  
OY 673 BOLKS-----RLDGKSGELGKRTTELDSKNQTLQKTEKYGCDONKLLTOOVXNLT 723  
Db 718 sqjaketeekllldeliraaesqakemkklrqqlaeseqklhleleknasekasaltr 777  
OY 724 EKKANVPKAYSEEMKSHOV--IYDOLAKKLSDVTHNYTEKLEMEKILMENASLSKNVSR 781  
Db 778 elqgrelklrlqenlasevagvketleklqllkekfaeas---eeavagvqrmqelvkn 834  
OY 782 LETVPIPERHEKEMMALKSNTTELKQULSKQSLNKKCGHDOEKISLMSB---NNDLAKT 837  
Db 835 l-----hqeceqfmnlseolekironlomeakftekdeeqqlakakelendleel 887  
OY 838 MSHOVVPVATHEIRIKTALSTLDKTNRELVDYKKCEDINOEPYKIDENELKRLNLENT 897  
Db 888 m-----kmsgdnsaqllkmlmedlrlkerdvaelqlklkanemasflqakledm 936  
OY 898 ONOVAKAYI---SLREHEE-----KASGLAKSMKQVONASBELAKTKKSOEELVTLH 947  
Db 937 --tvaeagqagqakahheekelerlledlekmetshmgqelkarverataetelkh 994  
OY 948 EELIADTIOECIKLKVAPITSLSECEKFKATEKELK-----EOLSOOTOKY 999  
Db 995 eellnqlnqlldt-----edklkgaateangllqelaelrlkqaeka 1036  
OY 1000 NTSE--BEAKKQKQENDKKELETLLOKDOKNVHINSEYTERALSRLK--TEELNRQK 1056  
Db 1037 kaaqaaedamngmqnteketelieslledtqlnaklqneldtlkennlkveelinks-k 1095  
OY 1057 DLL---OQYFAKEREKELVENAKOTSEILAAQTLLOKQHPLEOVESLKSLSGTIE 1112  
Db 1096 elltvenqmeefikelecl-kqaanaqagqisa--lqeenvkl--aeelqrs----- 1143  
OY 1113 TLKEELATVQKCYEBOQVTOLOKROMLENOKNSSVPLAEHLQVKEAPEKVGILKASLRE 1172  
Db 1144 --Tdevlshqkl--eeelravl-nqlllemkreesklkadeekasvqkslslsalille 1198  
OY 1173 KEEESQOMKTEEYSLOEIOHTKOALKELETRREYVDSLKYKATKSDLETOJSDLMEXLAN 1232  
Db 1199 kdaeletelnevtvlrgeneaeaksihavqtle-----sdxkklkelkvnl----- 1244  
OY 1233 LMKKEVECEFEVLAKKKELSAKDEKLHPF-----IEDEINDQERDCKSLTTITE 1285  
Db 1245 -----elqikenktrqjlaasagntdqdadeqeqelofinsvld 1286  
OY 1266 LOBRIOESAKOJETAOKDKITELLNDVBRKLONG 1320  
Db 1287 lqpkngqlkmkvemms-----eealing 1308

Search completed: April 10, 2002, 18:09:46  
Job time: 3830 sec





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2002, 18:08:47 : Search time 27.52 Seconds

(445,607 Million cell) updates/sec

Title: US-09-750-590-2.

Perfect score: 7039

Sequence: 1 MHSCWGCAPKRNQAAMWKN.....DYQALAIQIIRGLGWSA 1401

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2.6/pdata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/pdata/2/1aa/5B.COMB.pep:\*  
3: /cgn2.6/pdata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/pdata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/pdata/2/1aa/PCUS.COMB.pep:\*  
6: /cgn2.6/pdata/2/1aa/Backfiled.pep:\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB ID | Description      |
|------------|-------|-------------|--------|-------|------------------|
| 1          | 561   | 8.0         | 3248   | 1     | US-08-353-700-1  |
| 2          | 561   | 8.0         | 3248   | 1     | US-08-353-700-1  |
| 3          | 561   | 8.0         | 3248   | 1     | US-08-353-700-1  |
| 4          | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 5          | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 6          | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 7          | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 8          | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 9          | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 10         | 514   | 7.3         | 2101   | 1     | US-08-428-924-6  |
| 11         | 476   | 6.7         | 1312   | 2     | US-08-924-126-18 |
| 12         | 476   | 6.7         | 1312   | 2     | US-08-924-126-18 |
| 13         | 476   | 6.7         | 1312   | 2     | US-08-924-126-18 |
| 14         | 476   | 6.7         | 1312   | 2     | US-08-924-126-18 |
| 15         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 16         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 17         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 18         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 19         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 20         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 21         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 22         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 23         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 24         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 25         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 26         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |
| 27         | 466.5 | 6.3         | 1898   | 1     | US-08-056-200-94 |

|    |       |     |      |   |                 |
|----|-------|-----|------|---|-----------------|
| 28 | 371   | 5.3 | 1663 | 5 | US-08-353-700-1 |
| 29 | 346   | 4.9 | 1151 | 3 | US-08-840-006-5 |
| 30 | 325   | 4.5 | 1200 | 3 | US-08-840-006-5 |
| 31 | 325   | 4.5 | 1200 | 3 | US-08-840-006-5 |
| 32 | 318   | 4.5 | 900  | 2 | US-08-840-006-5 |
| 33 | 318   | 4.5 | 900  | 2 | US-08-840-006-5 |
| 34 | 318   | 4.5 | 900  | 2 | US-08-840-006-5 |
| 35 | 304.5 | 4.3 | 1065 | 4 | US-08-840-006-5 |
| 36 | 304   | 4.3 | 683  | 6 | US-08-840-006-5 |
| 37 | 302   | 4.3 | 1164 | 4 | US-08-840-006-5 |
| 38 | 302   | 4.3 | 1164 | 4 | US-08-840-006-5 |
| 39 | 302   | 4.3 | 1164 | 4 | US-08-840-006-5 |
| 40 | 297   | 4.3 | 1093 | 4 | US-08-840-006-5 |
| 41 | 296.5 | 4.2 | 1805 | 1 | US-08-840-006-5 |
| 42 | 296   | 4.2 | 955  | 1 | US-08-840-006-5 |
| 43 | 296   | 4.2 | 955  | 1 | US-08-840-006-5 |
| 44 | 296   | 4.2 | 955  | 1 | US-08-840-006-5 |
| 45 | 289   | 4.1 | 1104 | 4 | US-08-840-006-5 |

## ALIGNMENTS

RESULT 1  
US-08-353-700-1  
Sequence 1, Application US/08353700  
Patent No. 5599919  
GENERAL INFORMATION:  
APPLICANT: BLOOM, TIMOTHY J.  
INVENTOR: BLOOM, TIMOTHY J.  
TITLE OF INVENTION: NICOTIC ACID ENCODING A  
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DANN, DOREMAN, HERRELL, AND SKILLMAN  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
CREATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: RELEASE 11.0, Version 11.25  
CURRENT APPLICATION DATA: US/08/353, 700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELEPHONE: (215) 563-4400  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDNESS: single  
MOLECULE TYPE: linear  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
US-08-353-700-1

Query Match 8.0%, Score 561, DB 1, Length 3248:  
Best Local Similarity 21.94, Pct 126-219, Indels 514, Gaps 72,  
Matches 343, Conservative 245, Mismatches 465,

```

OY 251 ELAKO-- --BLOO--RULSOMLDRVNTXSNOPREHO-----NIDOLETEN----- 290
D 6 FEMEBOLPTITLOKIOELBGOULDKLAKKROKORFOLDSLEAPOKOTOKVHNKTECTNL 65
OY 291 -----EDLERKLRTOOBORILLDVNGLOLMEBVAVADOLESEKLSIL 339
D 66 KRENORIMCESLERTKOTKISHLEWESBOVBOJOLMSKROTEKIDOLEKMSSEL 125
OY 340 AMKKEOHESLARTALKSRFPYS-----DHLSGSH----- 374
D 126 ERSQOASAOVSLMNCPTPOKITPTPLTPSOVTSQSVYEDLKRYNKEVEERKLENAV 185
OY 375 -----KREDMLIKOGAMYTID-----SOCIST-----GMPVIMOSMSIRPLELALNOA 419
D 186 KALOAKKMSQTLQOATNHRDIAHRQASVSFSMOQERTYSILSSNSQOITKONEM--L 243
OY 420 SYSENELITKELEAKTPCSAKO-----DPLKON--ELAKVYAC-- 300
D 244 SYFSGBL-----EYTPRSOTLOKIDRANSFPFGSSPHLLDOLKAOBOLKINKHLELR 400
OY 460 -----KALALECEVKEVD-----SDQOTK-----OL 480
D 301 LOGHEKEMKOVNKFEBLOLLEKAYKELERKYLKNCBRLHDELVYTAQVASTRYAL 360
OY 481 EDALKOV-----OKRATES-----ISKYOMOTHE-----TALKER 511
D 361 EOKLAKITLEDLSOONNBARCSYDOKIKKEKEFEBELSHQORSPOTLOECTOKMAR 420
OY 512 LTRSD-- --ATGNHLEBELKIDOLKOMKYVEO-----ASAEVSKLNOIKONEM--L 558
D 421 LYOELDOAMNMHNVLOAELOLDTYKQOLENNLEEFKOKLACHAOAFOASOIKEMLERS 480
OY 559 VEERROESKLMENKRILOELSMCELEPERKQ-----RUTMEGOLK 602
D 481 MEEMKAKNNLAKSHS-- --QANRVCYHLELAKNITOCLOSONARAEKMAKNTSOTMUL 538
OY 603 DLSAKLALSLTPAKFPENKSLSLNDELAKKIL--DVEREYERSLNFTPRKLELENKA 661
D 539 DLOEKT-----NQOENSLTL-----EKLVLVADLEKORCSODLLKNEHLEBOLD 586
OY 662 KL-----AOHVKPERBOLKSHLEONS-----GELCKRITELTSKNOTLOKE 703
D 567 KLSKTESKALSLALELTKREBELK-----BEKTLRPSCKMSKREKLTDMESKLODSK 643
OY 704 IEXKULINKLITOOVYN-----UTTEKKNVPLKAKSPKKSIDYVODANKKLSIDVTH 756
D 644 IHN--LEOTLATOQIKSHENYERVTLEMDRENLSY--ETRMLINY-----LDSKSVEE- 694
OY 757 KYTEKLEKELKLMENALSKVNSRLTVEIIPREHEKEM--MALXSN-----TELKLO 809
D 695 -- --TOKLAYME-- --LOOAFES-----DQKHOLEIENKCKLSCTGOVYEDLEK 738
OY 810 LSLKLNK-----KCEID-----OKEXYSLSSEKN-----DLKKTSHOY- 842
D 729 LOLSNEIMDKDICTODILAEVESRLDLKSKOASVLNEDQNSLLAALDOQPMHNSFPA 798
OY 843 -----VPVKTHE-----ETKTLASSTLDTKNEBVLV 868
D 799 NIIGGOSVSEKREBLERLSDSPKSNALLONRVOLSEFSLSONOSOMDLOKCEBVLV 858
OY 869 VY-----KCEIDINOEPUKIDENLITKNELENT-- --ONQVAKAYVLSLREHEKMSGLB 919
D 859 IKGTEIEINLAKMDHOSFY-- --AETSORISKLODTSNHOQNVAVARTLSALENKEK-- --ELQ 914
OY 920 KSNKKVODNSAELLATYKRSQ-----BEIVTLHEBELAOKRELODTQECIKIKAJA11 972
D 915 LINDVTEQDAET-- --GELKKSNNILDSKLEKLOLSTLSTSLKDEKSSJISLNRKELEET 973
OY 973 SLEBEKCKKATEKELKQOLSQOTOK-- --WTSEPAKCKODNDIKLAKETILLKLODKXN 1031
D 974 OBNOLKLEINSLNOKCKMNLIOKSESNAVYIDEREKSSISELSOYKORLITLLOREBETG 1033

```

```

OY 1032 VHNESYETEBALSKTEELNROLDKLOLKYT-----BAKKREKLEVENAKOTSEILA-- 1085
D 1034 -----NAVYEDLSOKRYKAODERKMSLECLINBCTSLCEBNNRELEOLKEVAFKESHOEFTUL 1089
OY 1086 -- --KOTLOKONHPLVEYESLERS-----LSTLETLEKEE----- 1117
D 1090 AFMEENKONIMLELETVOALSLASSTDMONKNSBAGLOKLOLITLAKEDONKMOKEVNDL 1149
OY 1118 -----IKTORCYEXQOQVTOLOLROMENKNS--SYPLAELOVE-----AREE 1162
D 1150 LOENBOLKWKMTKHECONLSPEINNSVNEESERBONCKRONDELVEKESLSDSYNAO 1209
OY 1163 VGIKNSLRKEEESONKTYEVSLOSETONTR-----OALKLETRBYVDL----- 1209
D 1210 LWOELMILKRELEKLOESERKEKLOHLOLDTJRGLETSNLODMOSQOISGLCETIDNE 1269
OY 1210 SKRYAKSLELOTSID-----LHEKLANIKRYPERCEVYELHAKKLEKSKOKELH 1262
D 1270 EXYISOPHELSSTQNDNAHLQCSLOTPWKNLN--ELEKIC--ELLOAEKXELVY----- 1319
OY 1263 FSIODIRIOQERCDKSLTITTELQARLOESAKOIEAKNKITELLNDVERKQALNGLS 1322
D 1320 -----BLNDSRSEC-----ITATRNMAVEKQ-----LNLVY- 1353
OY 1323 QUTYQ-----SSQI-----SKROSQ-- --LDSLOOVYS 1348
D 1354 GLHAGLELEDOIPOGEFEPQEPNDQHPVSLATLDESNSVCHLTLSPKVOHMFALQOEKPLS 1413
OY 1349 LQOOLADADRQOEVIAIVRTHLLSLAAGHMD-----EDVOALLOITQ-- --M 1393
D 1414 LOSERHTILMDHOMV-----SSKMSLOTVOSLAKENVLSTNLNPODLKREMOQL 1468
OY 1394 ROGLVCS 1400
D 1465 EBOLVPS 1475

RESULT 2
PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Rep. Timothy J.
ATTORNEY: Rep. Timothy J.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
Expressed Kinase Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Damo, Dorfman, Herrell and Skillman
STREET: 1601 Market Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release 11.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
PRIORITY APPLICATION DATA:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janel E.
ADDRESS: 36 252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```



FILING DATE: 24-OCT-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/141,239  
 FILING DATE: 22-OCT-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 ADDRESS/DOCKET NUMBER: 1191  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2482 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-326-254-6

Query Match 7.7% Score 540: Db 1: Length 2482:  
 Best Local Similarity 21.7% Pred. No. 1.3e-21:  
 Matches 352: Conserved 291: Mismatches 540: Indels 440: Gaps 73:

Oy 33 DYKAVSLLAKGVNCKLDYKNSAHHVAVAKGNLECLMT-----LHSDUT 82  
 Db 670 ELEKICLQAEVLEYELTN-DNS-----ECITATIKAAEYGVKLLNRYVIL 717  
 Oy 83 TSDPTGRNALH-----LAAKYG-----HALCLQKLDYNCPTHEVDLGR-TALHDA 128  
 Db 718 NDS-----GLLNGELVEDI PCGPEFGQDHPVSLAPLDEN-SYELILSDKEVQHFA 776  
 Oy 129 AADCPSSIQ-----LLDHCASVNAKDVDCRT-----PLYATQKCRPTICQLLIR 173  
 Db 774 ELQKRPISQSEKILHDOCKSSKSLDTVYOSLAKENLYST----- 819  
 Oy 177 GADLSNDKQNRTA-LMLCEYCKDVAEVLKINKADVTLLDALGDHSYVYARIGDMLD 234  
 Db 820 ---NLKNFGCDLVKEMQLDEGLVSLSSCY--PSSSLSLG-DSSFYALLDQTE 872  
 Oy 235 LITLTK-----TASNSKNGELKMKGPSLQGRNL-----QMLDEVNTKS 275  
 Db 873 DMSLSMLGAVSNOCSDVDFCS--SDEENILTKETSAAPAKGYEELSLCEVYVQS 930  
 Oy 276 NQENHQTDLDEI-ENEDLEKELKIQOQRIILIDVNGLOQLNEFWVADLSEPERK 334  
 Db 931 LKRLKESMOSGLMKKEIQE--LEQLLSERQELDKQLYSEBQWQ--OKLTSVLE 987  
 Oy 335 LKSLIAKKEKHESLRTIETALKSFYFESDHLGSGHKKREKMLKOGQWYNTDSQCT 394  
 Db 988 MESKLAEMKQTDLSLELVAR-----LDLQ----- 1015  
 Oy 395 STCPVHMOSSMLR-PELALPNOASYENELKKELEAMR-----FCDSAK 443  
 Db 1016 ---LDLSRSLIGIDTEADATGR--NESCDSIKETSETTBTFRPKHDVHICDKDA 1067  
 Oy 444 DLKQLNELAHVAECKAL-ALEC-----ERKEDSDPQICQLE-----DA 483  
 Db 1068 QDLNLDLE--KLTETGALPTGDSGDSQPTNYEPGDEKTOGSSCELSISFSGKNA 1124  
 Oy 484 LKDY-----OKRMYESEGKVKOMQHLLAKKEHLESDAQCNRIMELKQDLKDKMYK 538  
 Db 1125 LVPMDFQGNODEIHNLQKRETSNEMLL-L-HVTD-----NDRVSESLNEMKEID-- 1176  
 Oy 539 EGASAEVGLKRLNQTQENELV--EPRDRCKLMKERRKQLELSMCELEKERRKRLT 595  
 Db 1177 -----SKRLHDEVLQKTIKACLEIKTKVGLKKEKSDSKLETFSCDHQDLTPK 1229  
 Oy 596 EMGQGLDLSAKLALSTPAKFPNNKSLSLNLEAK-KLVDYREYRSINTEQPLK 654  
 Db 1230 TSBGINSDELPMHAKSSREDIGDNA-----KVMSDEKRPDLVDENELSRISSEKASIEH 1284

Oy 655 ELEKIAKLIAOVHVPREHBOUKSRLEQSGELGRITELTEKNOFLQKEIEKVCLOKRL- 713  
 Db 1288 EALYLEAOL-----EYVOTEKLCLEK-----DNEKNQKQV--VCLSEBELS 1322  
 Oy 714 LTQVNNLTTPKAVPLK-----VSEKKSQHDVIVDILNKRSLSDVTKYKTEKLE 766  
 Db 1323 VYSEKNOGLBLOKSKTETATLDDOSKTKER-----TQELSHSGELHLCYQAVAKR 1378  
 Oy 767 KILMEKASIKVNSHLPVLPPEPR-----EKPMALKSNITELKQLEKNTK-- 815  
 Db 1379 EKTELQTLSDVSEL-----LKQKTHLOEKLSQKSOALSTKCELENOJLANKKE 1434  
 Oy 816 -----KQGEDQKITYSLSENNOLKRTMSHOVY 843  
 Db 1435 LKVESSELOARLSSEYERKLNKALPAALVKGESALRSTQEBOLHNRIGELV 1494  
 Oy 844 DYKTHEELKASSTLDKNEELDYVQKKEQI-----NRPVYIKDN-----RI 889  
 Db 1495 RLEADKQOHIAKRLKEREENDSLKIKYENLRELOKSENGEVELYLDASNAEVE 1554  
 Oy 890 LKRLLETONOVYA---EYISLREHEKMGKRLKMKKYVONSSE-----LAKYK-- 937  
 Db 1555 LKTOIEBVAHSIKVELDVLTRSEKEMLT-----KOIOEQGQULSELKLLSPSL 1608  
 Oy 938 -KQGEIVLTUELIAQKRELDTQECITKAPILYSLCEKRYKATKREKLEKQISQ 995  
 Db 1609 EKKQDAEL-QIKKE--SKTAVEMLOQLMEVNAVALCDQDEIKVATQSDLPPIEB 1664  
 Oy 996 TQKYNPSE-----EAKKQKQENDKLKEKITLQKDL-----KDN 1031  
 Db 1665 IOLNSTIKRALALEDEKQOLVLOQLKSEHMAHLDGAVENLEKLELARTNOEHA 1724  
 Oy 1032 VALNSETYERBALSKTEBELNROLQLOKTEKAKKEXUYENAKQTESI-----L 1084  
 Db 1725 LEAPNSKEVETLAKITEGMSQKLELQVYTIKSEKNIENLEKQSRISLEITINS 1784  
 Oy 1085 AAGTLLQKQVHPLEQVESLKSLSGTYITLKEKLE-----TKQCKYERQOQYAT 1133  
 Db 1785 SFENILOKRE--QEKVOKERS-STAMEMLOQLKELNRYVALHDOEAKKQNTSS 1841  
 Oy 1134 -----QLRKLEKNSSVPLAELQVKEKPEKEVIGIKASLEKRESEKQNTSE 1183  
 Db 1842 QVECELEKQQLQGLQDEKANNYIVAGSNV-----GLIO-EVEDKQKLEKDE 1891  
 Oy 1184 VSKIOSEPTQKQALKLETRREVDSLKYKATYSDLTQI--SDLNKJLANKKEEYVE 1242  
 Db 1892 ISRLKQLODOEOLVSKL-----SOVGSQDLKQKQENLEKRLNYLE---Q 1935  
 Oy 1243 EYLHAKKELSANDEKLHPSS--IEQERK-----DOO--EKDCSLTITTELORIE 1292  
 Db 1936 KLYQDSKNSLSQDTLQVYKLSYKMLELDELKMKDMSFVAKYKMTAEETLOREHIE 1995  
 Oy 1293 SA-----KOIPEAKDKIT-ELNIVYRPLKQALNGLSQIYVSGSGSPSKR-----QSQ 1337  
 Db 1996 MACTALEQELSESGKRLAGLOVLLLEIKSKQDKLELTLNLSLTKSLSDCMKQV 2055  
 Oy 1338 LIDSLOOVRSLOQLOADADROHCV-----LAIYTHLLSAQGHMDEVOA 1386  
 Db 2056 KEGVYMEELATQYOLMBKRNKQNALLDPTNQYEVLOTYKELKLSKESQKLEL 2115  
 Oy 1387 LLD 1389  
 Db 2116 LK 2118

RESULT 4  
 US-08-466390-4  
 Patent No. 5666390  
 Application US/08466390  
 GENERAL INFORMATION:  
 APPLICANT: TOUNART, GARY  
 APPLICANT: LINDARD, GRAHAM P  
 TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

```

: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HIRNITZ & THIBEAULT
: STREET: BOX 225 HIGH STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, version #1.25
: CURRENT PATENT DATA:
: APPLICATION NUMBER: 05/08/466,390
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ. EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MIP-013
: TELEPHONE: (617) 248-7100
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ. ID NO.: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MODIFIED TYPE: protein
: US-08-466-390-4

```

Query Match 7.3%; Score 514; DB 1; Length 2101;  
 Best Local Similarity 20.1%; Pred. No. 2,96-20;

Matches 309; Conservative 274; Mismatches 524; Indels 434; Gaps 61;

```

OY 185 KONTATMACEYCKRAVEV-LTKNGADVTLDDAGHDSYVARKGONLDTLTLLTKAS 243
Db 215 QNRRLKOLADERSNDELDELAKNRKLLTKENKA-----QIAMQOKRIDRLALNEKO 268
OY 244 RNS---NKGRELAKKQPSLOOR---NISQMLDENVTKRSKORHONQIOLLEINENDKR 296
Db 269 AASRLPEKLEELRDKNESLTKRLHETLKOCOD-LKTKESQMD-RKINQLEENDDLSRK 326
OY 297 LAKTOOPORILLDKVNL-----QDLMEVYVADDESEK--EKIKSLIA 340
Db 327 LKRAASHLOLOLALNELTEHSAQOMLEKQALKEKELSA--LQDKRKLEENKIILO 384
OY 341 AKKQHESSLRTIE-----ALKSHRYTESDHR-- 367
Db 385 GLTSOLEHLSLOLDNPPCKGKGVLDVQLQETLKQSAATLAANTQOLANQVEMLETKNG 444
OY 368 -----LSSGSHREKEMILAKO-----COMPTD 390
Db 445 QOEKALLKRGHREKQOLESLLTDLOSSISNLQAKKELEBOASQAGKATLAQVAVSLT 504
OY 391 SOCSQSPYHMOSSMKAPLELALPNQASTSEMLKKELEAMRTFCDASQKORLILAN 450
Db 505 SELTLTANTTQODDELATLQQAQKKAQALQAT--LQQAQSQSLKQVQED--LSS 558
OY 451 ELIAKVAACAKALALCERKVEKSDPO-----IQOLEDLALQDKRKRYTESBQYK 499
Db 559 SLIKKQQLKEVAEKQATRDQAHQOLATTAAREPRALEKRDALIKOLEALEKEKAAKLE 618
OY 500 QKQHTLALKEHLTSDATGN-----HRLMEELKQILKDKKRYTESASAEVGL 548
Db 619 ILOOQLOVANE--ARDSAQTSYQADQREKALEKSKVEVEELQACVETRAQDEHONQAVEL 676
OY 549 RNQIKQNMILVEEPR--DEGKLMEKRIKLLQELMSCK--LREKRGKRLTEMGQILD 603
Db 677 ELQHSREQDAKTEREVAQEKQDLOLOLAKESIKVTKYTGSLSEBER--RAADALEQDRC 735

```

```

OY 604 LSAALALSTPVEKFNKSLLSNELNEKAKKLDIVEREYKSLN-----ETRP 651
Db 736 IS-----ELAKE-----TRSLV--DQKHNRKLELEBQKGLANILQJLGAHQALTEV 784
OY 652 LKRELNLKATLQAVKPREHEBQKLSRL-----DQSGKQKRIETLISNTOLO 701
Db 785 LRRLT--LEMAAHOHTASECEQDLYVEAARQGYECSQEEPAQYCAMOE--QLTMTK 839
OY 702 KTEFYQJONKLLTQVNNLTETKRVPLKAVSEKSKSHDYIVQDINKKLSDTHTAKTER 761
Db 840 EECRAKROLDKAKRVAAGSHSE--LQTSQDQKN-----LAEILHANLALQVQOK 891
OY 762 KLEMEKL-----LMEH-ASLSKYSRLTEVF-----TPER 791
Db 892 EVRAKQALDLSLTQEMKAATSKREVRLETLTKKAGQOETASRELYVEDARAGDQRP 951
OY 792 HEKE-----MAUKSNITELKQSLSEL-----NKGGEQOKSYLMS 829
Db 952 LEEQDQRPQSTQALQDMEKRAVQNGNELEIRALALMESQGOQOEQOQREKVALILO 1011
OY 830 ENN-----DLKTHQOYVNTVHTKTKTALSSLTQKTNELVDYK- 870
Db 1012 EFGHAQADLALAKARALEKRILONLALMDQREVPATLOALALHATLEKQKQDLANTNG 1071
OY 871 -----KRCEDINQEFYKIKDENELTKRNLNTQNO-----VAAEY 905
Db 1072 LEMAQILKLELKQTVKOLKE--QIARKEREHASQSGASQAPAGRTERTPKALPAAEY 1129
OY 906 ISL-----REHEFKMSGLKSKMYQDNSEF-----ILAKTKSQDEVTLHEBIAOK 954
Db 1130 SKLEBOCQKQKQADQADSLERSLEAFASRAEROSALETQQLDEKMAQILHQSALMSQ 1189
OY 955 RELDTIOECT-----KLKAPITISLEB--CEBKFKATE--K 986
Db 1190 HELMAFKRKQVODHSAKDEMKRAQVANGKQARAKKSLISLSEFVSLINROYLKEDESK 1249
OY 987 ELKQOLSOQPKYNTSEBAKKQKQNDKILKKEILLQ-----KDKRKWHTIE 1035
Db 1250 ELKRLVAMESEKSKLESCACRCQHPQATVPELONNALGLGRCRQASGAEKQVAVSE 1309
OY 1036 NSGETRALSRTTELNRDLQLOVTEAKKEKEL--VEENAKO--TSEIILAQTLIO 1091
Db 1310 NLQBELTQSAEBAEELQOBLAMQKFFQOEQALSTLOLQETSTQVALSELPRKHLICQ 1369
OY 1092 -----KOIYPLQVNSLKSLSGTEITLKELEKTKOR-----CYREQOVT 1132
Db 1370 LQMGQAAKRRH--REELQKQKQAG--LAMELIRQNSLDELPLRQNVKALQDEKRA 1423
OY 1133 TQLR-----QMLEKQKNSVPLAEHL-----QVNEKAFKXVGI 1166
Db 1424 QOURLREKASYAOLSMYLLKKAAGLLEBKRGDLERANIGROFLEVLQDAEEYVOELAV 1483
OY 1167 KAS-----LREKESQONKTEESKLSQSEIONTQKALKLLETRVYDLSKRYKATISPLETO 1222
Db 1484 RQADATRLALFYQRBQSTAFLELEVTAKTATQAK--VKVLEBPO-----ROSEBQKTLAQ 1536
OY 1223 ISQLEMKTLALRRYCEVCEVPLAKKKEKLS--DKEKELTFSIE----- 1266
Db 1537 VEELSKTIL--S--SDQASVQOQKILKAVQKQSGSQEPQAFQADQMLQNELOALSQKQ 1591
OY 1267 --QELKQDQERDQSLTTITLQELRQIESQKQIE--ANDKNTIELLVNDRYKQALMGLS 1322
Db 1592 AAKHTIKQEKKAKTKHTQDAKQONQOELQDQUSLSQLOLEK--ELRAVEARLCHIELQ-- 1646
OY 1323 QLTYSQSPKQKQSLLSIQOQYVSLISQLOLADQNDQPOYQ 1363
Db 1647 ----QAGIKTKBAQTCYGNHLTQAVESLEQVAVHQADQADQL 1663

```

RESULT 5  
 US-08-470-950-4  
 ; Sequence 4, Application US/08470950

```

1 Patent No. 5698439
2 GENERAL INFORMATION:
3 APPLICANT: TOLUATLY, GARY
4 ATTORNEY: JAMES D. P
5 TITLE OF INVENTION: NOVEL NUCLEOTIDE CELL TYPE MARKERS OF THE
6 NUMBER OF SEQUENCES: 6
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: TESTA, HORWITZ & TRIBAULT
9 STREET: 125 HIGH STREET
10 CITY: BOSTON
11 STATE: MA USA
12 ZIP: 02110
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patentin Release #1.0, Version #1.25
17 CURRENT APPLICATION NUMBER: US/08/470,950
18 PENDING APPLICATION NUMBER: US/08/470,950
19 CLASSIFICATION: 536
20 ATTORNEY/AGENT INFORMATION:
21 NAME: PITCHER ESQ, EDWARD R
22 REGISTRATION NUMBER: 27,829
23 REFERENCE/DOCKET NUMBER: RFP-013
24 TELECOMMUNICATIONS NUMBER: 617/248-7000
25 TELEFAX: (617) 248-7100
26 INFORMATION FOR SEQ ID NO: 4:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 2101 amino acids
29 TYPE: amino acid
30 TOPOLOGY: linear
31 MOLECULE TYPE: protein
32 US-08-470,950-4
33
34 Query Match 7.38; score 514; DB 1; Length 2101;
35 Best Local Similarity 20.18; Pred. No. 2,96-20;
36 Matches 309; Conservative 274; Mismatches 524; Indels 434; Gaps 61;
37
38 Db 185 KONTRALMISERCECKDNEY-LIKNADPTLLDAGHDSYATIDDDTLITLKAS 243
39 215 QMRILKQADBSNDELDELLENKRLTEND-----QAMQORIRLALLNEKQ 268
40 244 ENS-----NGRELLKKRGPISLQDQ--MLSQMLDEVTKSRQPHQIODELENEDELEK 296
41 269 ASPLREKRELEMLDNESLTKMLHETLLKQCD-LATERSQMD-RKTNGLSENGSLSEFK 326
42 297 LKRTQDQRIILLKNGCL-----QQLNEEYVAVADPSSEK-EKLSKLLA 340
43 327 LRFPSLQIQDQDALNELTEHRSKATQEMLEKQALEKESLAA-LQDKKLSERKNEKLT 384
44 341 AKRKQHSERLRTIE-----LMSRKPFPSDH- 367
45 368 -----LGSSEFKKEDMLKQ-----GQYMTD 390
46 445 GQFALLAKRHFPEEKQOQLSLITDQSSLSNLSQAKLEBDQASQNHGRLTQVASTL 504
47 391 SGGTSCPMVHNGSRSLRPLETALRYPNQASVSENETIKKELAMRTTCSAQRDLKON 450
48 505 SETLLVMTATQODQDLAGLKQOAKERQADLAQT-LQDDQASQGLRQVQD-----LSS 558
49 451 ELAHVAVESKALALECEPVKEDSDQ-----IKOLEDAKLVQKRMTESEGVK 499
50 559 SLKQVQQLLEKVAQDQATPDQADQALTALEPESASLEKQDLYLAKLEKKAALDE 618
51 500 QKTFPLAKLEHTLSAATG-----HMLKEELKQKQKMYUYEQAASAVGKL 548

```

```

Db 619 ILQDDQVQANE--ARDSQSTVQADQBRKELSKVELEQACVETANQDEQHQADQVEL 676
549 RNIQKONHUYEKKR---DEKLEENKRLKQLEKSE--LEKRGKRLTPEHQGLD 603
677 ELALREDOKATENEKVAQDEQDQDQDQDQLEKSLAYTKGSLSEKKR-RAALALEQDQK 735
604 LSKALSLAIPAKPENKSLLSNELEKAKKLLIVEREYSLN-----ETRP 651
736 IS-----ELAAE--TSELV--DQKREKLEBBERAGRGELRLLQSGAQAQVEY 784
652 LKRELEMLAKKLAQHYVPEHEQDLSRL-----KQSEKQKITHLETQKNQTLQ 701
785 LRRL--LEMAAQHTASECEQDQVEVAMQDQEDQSQEQAQGAPE--QMLTKL 839
702 KELEKVCIDKRLTQGYNNLTTEKKNVLEKVEEMKSHDYVDLKNKLSQVTKETK 761
840 ECKKRAQDQENKAKQALSHSE--LQDSROK--LAEIHNKLAQADQVQK 891
762 KLEMEU-----LMEN-ASLSKNGVLETFV-----TPPR 791
892 EVRRQGLADLSTLQERMAKLSKVEVRLLETVRKAGQOETASRELYEPARADDPQ 951
792 HEK-----MVALKSNITELKQSLSE-----NKKSGDEKITSLS 829
952 LEEQQRQFCSQVQALLQDNEBQDQCNELERKALAMSQOQDQEDQDQREYQALQ 1011
830 EKN-----DLKTNHQYQPVYVTEHETITALSSTLDKTRRELYDV- 870
1012 ECKRAQDQALEKAAKRAELERLQNALREQREVPATQJALALATEKCKQDQAKLNG 1071
871 -----KCEJDINQEFVKTQENELTKSNLENTQO-----VATY 905
1072 LEAAQKLELEBLQVYQKALB--QLKKEKHSQSGQSDQAGRTPTQKALALAV 1129
906 ISL-----BEHEKMGSLKRSKHYQDQNSAE-----ILAKYKSGOEITVHEBIAQ 954
1130 SKLEQCCQKQDQADQSLERSLEBARASBQESALLETQLQDQLEKQALSHQSASLMO 1189
955 REQDTQDCTI-----KIKYAPITSLSE--CERRKATE--K 986
1190 RELAAFRKQVDRKSAQDEKKAQVANGQAKENKSLLSSEBSEYILMQVLEKESK 1249
987 ELKQDSQDQGYKNTSEKAKKQKQENQKQKXKRETLQO-----KDLKDRNVITE 1035
1230 ELKELVWESKQKLESCQSCQDQVQVYVPLQNALLCGRKCSAGREKQVASE 1309
1036 NQVETRALSKRETELRLQKDLQKQVTEAKKEREK-VEENAKQ--TSEILAQTLQ- 1091
1310 NIKREPLTSQARBAELQDQELKQKQKQKQASLQDQHTSTQVQVLSBELKHNKQD 1369
1092 -----KQHPVLEQVSKLSLSTLETLEKLETKQK-----CYEBOQVY 1132
1370 LQADQAAAKRH--RELEDSQKQAAQG--LBAELRLQRELEDLPLRYKQAVDGRFA 1423
1133 TQLR-----QMLENQKNSVPLAEHL-----QVKAPEVGI 1166
1424 QADPAKRASTADQMSLKRNQALNQLGSRANLQDQPLEYELDQANRKYQVQELANV 1483
1167 KAS-----LPEKEREQKNTPEVQSLQSEIQNTQVQKLEKTEVEVULDSYKATKRDLEQ 1222
1484 RADEKPLAVQREQAQSTARELEVWIAKKEAR--VKVLEEQ--RQDERQKQLQD 1536
1223 ISQNLKRLANIKKYEVEVCEVLIQAKKELSA--KQKELHLSIE- 1266
1537 VEELSKLAD-----SQDASKVQDQKKAQVQDQSQDQADQADQVQALQDQALQSKQD 1591
1267 --QQLQDQKQCSKSLITTELEQRITQSAQOTE--AMDKITELLYLNDVRELYQKNGLS 1322
1592 AAEHTQKQLEKATNTYQAKQKQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 1646
1323 QLTGSSPSPRSQSLDSIQDQVRSYLQDQALDQADQVQ 1363
1647 -----QAGLTKREAVQYCHNTQVQVRSLEQVQVAAHQADQDL 1683

```



RESULT 6  
US-08-467-781-4  
US-08-467-781-4  
Patent No. 5,780,594  
GENERAL INFORMATION:  
APPLICANT: TONKATY, GARY  
APPLICANT: LIDCARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
CORRESPONDENCE ADDRESSES: 6  
ADDRESSES: 25 EIGHT STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: RASNET Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: Release #1.0, Version #1.25  
APPLICATION NUMBER: US/08/467,781  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESO, EDMUND R  
REGISTRATION NUMBER: 27,829  
TELEPHONE: (617) 248-7000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2101 amino acids  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-467-781-4

Query Match 7.38; Score 514; DB 1; Length 2101;  
Best Local Similarity 20.18; Pred. No. 2,9e-20;  
Matches 309; Conservations 274; Mismatches 524; Indels 434; Gaps 61;  
189 KONTAKALMCCGCGDANVY-LIKNGADVTLLDAGHSSYVARIGDNDITLTILKTAS 243  
DB 215 OMRLKQKOLADERSNDELLELAENRILITKXK-----QTAMQORTIDMLLNKKO 268  
OY 244 ENS-----NKGRLLKKKGPLOOR--NLSQMLDEVTKSNORHONODLEENHETLR 26  
DB 269 AASPEPELELELDKKNRESLITMLITLTKOCOD-LATKESQMD-RKINOLSENGSLSPK 326  
OY 297 LRATQOQRLLIDLVNKL-----QQLNEBYVAVDDLESK--EKLASILA 340  
DB 327 LRFPASHLQDLDALNEITRENSKATODMLFKQOLEKSLNLA--LDKPKCLEKKEKITLQ 384  
OY 341 AKKKEHSESLRTIE-----ALSRKKYFESDH- 367  
DB 385 GKLSQLESHLSQLDNPPRQEKGEVLGVLTLETLKQEAATLANNTOLOARVEMLTETRG 444  
OY 368 -----LGSNFKPKEDMLTKO-----GQMYTFD 390  
DB 445 QOQAKLLAERHGFEEKQGLSSLLIDYDSSINLSQAKELDQASQAKNARLTNAVASLT 504  
OY 391 SOSTSTOMVHNGSRNMLRPLELALPNOASVSENMETLKEELAMETCSAKAKORILKON 450  
DB 505 SETTLVLAITQDQDELGLAKQOAKQKQOLALQTF-LQOQDQASQGLRHQVQO-----LSS 558  
OY 431 ELAHVVECKALALCEKVEKESDQO-----TKOLEBALDKVCKRRESSEGVK 499

DB 559 SLAKQKQDLKVEAKQKQATRODHADQDLTAABERBASIREMDALKOLELSEKKAAYKLE 618  
OY 500 QMOETIRLALKEHDSYDAKQ-----IRKLELDKOLKQMKVYTCGASVNGKL 548  
DB 619 ILQOQLOVANE--ARSDAOTSVTQAQNRKALSKVEBLQACVETANQDNEHQAQVAVEL 676  
OY 549 RNOIKONMIVEEKKR--DECKLVEEKRLQKELASNC--LREKRGRLNREBOGLKD 603  
DB 677 ELLJSEBOKATEKEREVAQEDQDQDQALQKESLKYVKGSLSEKKR--RADDALEQDRC 735  
OY 604 LSKALALSTAKREKNSKLSLWELEKAKKILVEREYSUN-----ETRP 651  
DB 736 IS-----ELKAE-----TRSLV--EGRKREKLEERAKNGKLELALLOLSEANQAEV 784  
OY 652 LKRELENLKAKILAOVKKPREHOLKSLR-----EOKSEIGKRTITELTSKORTLO 701  
DB 785 LRREL--AEAMAAQHTASEEDOLKVEAMHNDGYEDSOEAGQVAMQOE--QMLTK 839  
OY 702 KELEVCLDKMLTLQOVNKLITKMKVPLKVEBKKSHIVYIDLKLSVDYHNYEK 761  
DB 840 EEEKAKQLOLQAKENYAKENISHSE--LQJSHQOK--LALHMLANALQDVER 891  
OY 763 KLMEKLT-----LMEN-ASLSNYSNLEVEY-----LPRP 791  
DB 892 EYBAQKRLADQLSTQEKMAATSEVARTLETIVRAGQEOETASRELVEKPAKRDQRPEN 951  
OY 792 HEKE-----MAALSINTELKQILSEL-----NKKQEOEKYIVSLMS 829  
DB 952 LERQDQRCQSTQALQAMERERDQKQNELERLMAALMSQDQOORQOQREVARLTO 1011  
OY 830 ENN-----DLKTSHQVYVYKTHEELKALSTLTKNGLYK- 870  
DB 1012 ERGRADALALEKAAKALEMYRLQALNMQVREPATLOGLAHALFEKQKIDQLKRLG 1071  
OY 871 -----KKCEDINQEVATIDENELIKRLNLTQNO-----VKKEV 905  
DB 1072 LEMAQILDEELNRYVQKQLE--QALAKREKMSGQASQASQAGRTPTGKLEALREV 1129  
OY 906 ISL-----REHEKMSGLKRSKMKKYQDQNSAE-----LAKYKKSQEPYLYLIELAK 954  
DB 1130 SKLDQOQOQOQDQDLSRLSLAERASBRDQALFTLOQLEKQALGELGQSLMSAQ 1189  
OY 935 REIDTQECI-----KIKVAPITSLSE--CEKFFATSE--K 986  
DB 1130 RELATFTYVQDSKAKDEKQAKQVAKQOEAERKNSLISLSEVSTILKRVLEKESK 1249  
OY 987 ELKQDLSQOTQKYNSEEEKKQKQDQDKLKITLTLQ-----KLQKNNH 1035  
DB 1250 ELKRLVYMAKSEKQKLESCACCPQRPATVPELOMALLCSRCNASSRBAKQVASE 1309  
OY 1036 NSYETERALSRTTEBYNRQDLQKQATLAKKREKL--VEENAKQ--TSEILAQOTLLQ- 1091  
DB 1310 NNRQDLSQAEKRELDQELKAMQEKQKQKQALSTQLEHNTSYQALVSELYLANNLSQO 1369  
OY 1092 -----KQVRLQDQVSEKSLSGTITLQKAEELKQV-----CEKQDQVY 1132  
DB 1370 LQASQAAKRN--RELEDSQQAQAS--LADILQVRLEDELYRLQKQAKQABRTH 1423  
OY 1133 TQLR-----QKLENOKNSVPLAENL-----QVKADEKVEGII 1166  
DB 1424 QDLQAKESVADQELNKLKAKHQLAENRNGLRANGLRQFLEVELDQARPKQVETALAV 1483  
OY 1167 KAS-----LREKEESQKTEVSKQSELOMPTQAKKILTRREVQSLKTKATKSQDLEQ 1222  
DB 1484 RAEDEFRLAVQREAGOSTABELEVYAKYBAK--VAVLEERO-----RODERQKILQD 1356  
OY 1223 ISDLNKRLANLNKRYEEVECEVLIANKKELSA--KQEKELLHNSIE----- 1266  
DB 1357 TQKESKTLAD-----SQQASKYQDQALQAVQAGQESQDQARQVQDQNLQDQALQSOQK 1591  
OY 1267 --GEIRDQORQDKSLTITTELQRIQSSAKQIE--AKNKITTELLANDVERKQALNGLS 1322

Db 1592 AAEHYLQMEKAKTHYDAKQKQNOELQRLSELOLEK--ELRAERELGHELO--- 1646  
 QY 1323 QUTYGGSPRSRQSLDLSQOQYBRLQOQDADARQHEV 1363  
 Db 1647 ----QACLKTEKREOTGRHLTVQVRSLSAQVAHADQOQDL 1683  
 RESULT  
 US-08-483-924-4  
 Sequence 4, Application US/08483924  
 Patent No. 589370  
 GENERAL INFORMATION:  
 APPLICANT: TOUKATY, GARY  
 APPLICANT: LIDCARD, GRAHAM P  
 TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 STREET: 125 HIGH STREET  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 OPERATING SYSTEM: GENOSIS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/483,924  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: FITCHER, JES, EDWARD R  
 ADDRESS: 27 829  
 REFERENCE/DOCKET NUMBER: WFP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2101 amino acids  
 TYPE: amino acid  
 TOPOLOGY: 1  
 MOLECULE TYPE: protein  
 US-08-483-924-4

Query Match 7.38; Score 514; DB 2; Length 2101;  
 Best Local Similarity 20.1%; Prod. No. 2-96-20; Indels 434; Gaps 61;  
 Matches 309; Conservative 274; Mismatches 524;  
 165 KQNRALALGCGYCCQDAVEY-LIKNAQVUTLDAIGHDSYVAHIDMLDLITLKAS 243  
 Db 215 QMRILKQJLADERSNDELELELAENKILITBKA-----QVAMQORILALLINKQ 268  
 QY 244 ENS---NKGRELKKQPSLQOR--NLQSLMDEVTKSNQREHONTQDLETENEDAKR 296  
 Db 269 AASPLEPELELELDKKNLSLTKMLITETLAKQCD-LTKESQMP-RKTNQSLSENGLEFK 326  
 QY 297 LKRLQQRRLILDKYGL-----QQLNEEWVADLTSEK--ELKLSLLA 340  
 Db 327 LRFPSNIDQLDADALMELTLEHNSKATQEMLEKQALEKELSA--LDKKCLEFEKHEILO 384  
 QY 341 AKRKHESLRTIE-----ALSRFYKFSQH- 367  
 Db 385 GKLSLEPHLSQLDQNPQKGEVYLDVLTETLKEGAPLANNTQALNNTQLEMLTEFG 444  
 QY 368 -----LCSGSPFKKEDMLKQ-----GQMYMTD 390  
 Db 445 QGQAKLTLRNGHFEERKQSLSLTLDQSSISNLQAKLELQASQAHGRLTQVQASLT 504

QY 391 SQGTSQCPVHQSRSMLRPLELALPNOASTSENEITLKELEMTFECDSAKQDRILKQ 450  
 Db 505 SELTLNATLQOQOQELAGLKQQAERKQALQAT--LQOQDQASQGLRHQVEO---LSS 558  
 QY 451 ELAHRVAECKALALECEKVEDSDQ-----IKQLEDALKYQAKMTSESGK 499  
 Db 559 SLQKQKQLELVAKQKATNQDIAQDQALNABERFASLRERDQALQALAEKREKKA 618  
 QY 500 QMOTFALKEHLTSDPAATG-----HYLMEQIKQDKQDKYVEBSAASVLT 548  
 Db 619 TLQOQVQANE--ARQSQTSYQQAQEKAKELSKRYEELQACVETARQDQDAQV 676  
 QY 549 RNOIKQNBILVEFR--DEGRILMEENKQKQELSMCE--LEKKGKRLTPEQQLQ 603  
 Db 677 EQLSSQOKATKEREKADQENQDQOQDQSLQELNAYTKGSLJEERK-RADMLLEQQR 735  
 QY 604 LSKAKLALSTPAKPEKNNKSLSMELMEKAKQILVEEYERSYL-----ETP 651  
 Db 736 IS-----ELAAE---TRSLV--EQKREKRLDEERAGRGLELRLLQAGAHQV 784  
 QY 652 LKLELENLKAKIQAQHYRPEHEQLSRL-----EQNSQELGRITETYSKQTL 701  
 Db 785 LRREL--NEMQAQVHSECEQDQVEKNAHQDYEDSQENQACMPQ---QMLTK 839  
 QY 702 KELEKQCDMLKLTQOYNLTETTKKYNVLVESEKSHVDYVDQNKLSQYVNHUYEK 761  
 Db 840 EECERKQELQAKKQVACISHS--LQISQONK---LAEILNANLALQAOYQK 891  
 QY 762 KLEKEL-----LMEN-ASLSKNVSLTFV-----IPPER 791  
 Db 892 EYRQKQLADDLSTQERMAKASKEVAKLETILVRKAGQDQGTASRELVEPRADQDR 951  
 QY 792 EKKK-----YMALEKNTTELKQSLSE-----KKKQSDQKITYSLMS 829  
 Db 952 LEQOQRFQSTQALQAMERELQDQKMELELYRQALMESQOQOQEMQOQREYVATL 1011  
 QY 830 ENK-----DLKTNHQYVYKTHEEINTALSTLDTNRELDVVK- 870  
 Db 1012 ENRQAQDQALAEKRAELENKILQNLNEDQREYRATLDQALANALTEKQKQELAM 1071  
 QY 871 -----KKQCDINQEPKIQDQENITLAKRNLQNTQNO-----VAYE 905  
 Db 1072 LEAQIKLELELRQTVQOLKE--QAKKEKEHNASGQASQAPAGERTQPKLEALAEV 1129  
 QY 906 ISL-----REHEKMSGLKSNKKQVQNSAV-----TLATYKQSOEIVTLIEIA 954  
 Db 1130 SKLSDQCKQOQDQSLNSLELARSASRESDSLQTLQOQLEKQKQELQSHQSLNS 1189  
 QY 955 RFLDTQECI-----KLKAPITISLEF-----CEKKFATE---K 986  
 Db 1190 RFLARFQYQDHSKADQEMQAQVAGQARPKNSLISLEEVYSLNQVULEEDEEK 1249  
 QY 987 ELKQSLQOQTKNTYSEBKSKQSNKIKKQITLTLQ-----DKLDKQNVHE 1035  
 Db 1250 ELKTLVMESEKQKLEBSCQCHQOQVAPVLTQNALMLGSRKASQAGREKAKORAYE 1309  
 QY 1036 NSYETFRALSKYTELRQKQDLOLTQETKAKKELK-VEENAKQ--TSEILANQD 1091  
 Db 1310 NMRQELTQARERELQDQELKMAQKQEFQKQALSTQLEHTSTQVQVSEILPANNQ 1369  
 QY 1092 -----KQHVPLEQVSKLSQSTLETLKELTQK-----SEYEDQOYV 1132  
 Db 1370 LQADQAAAEKRN--REELDEQSKQAAG--LQALVLRQARQELPRLQVAKQDQ 1423  
 QY 1133 TQLR-----QMLENQKNSVPLAEHL-----QVEMAKREYGI 1166  
 Db 1424 QDLRAEKASVYAEQSLMKKNGLLABERNGLSGRANGLQRLVELELDQREKQVQELAAV 1483  
 QY 1167 KAS-----LBEKREBQKTEVEYSKQSELTQMTQAKLLEKTEVEVULDSYUATK 1222  
 Db 1484 RADQETYLEVQREQAQSTARELEVMYAKQKQK--VYUULEEEO---RQDERQKLTQ 1536  
 QY 1223 ISDLNRKLANIKRYEEVCEVYLAKKKELSA---KQKELHLPSTIE----- 1266

```

Db 1337 VEELSKRLD-----SDASRVOOQKILAVNOAGSGSOQAPROQADLELOAOISOKEO 1591
Oy 1267 -OETKQOEDCKSLITTELOPOTLESANQOE--ADNDKTELLINDVETLOKALNGIS 1322
Db 1592 ALENKILQNEKATKTHYDAKKNQOQOEDLOPOLSLEDOLEKEM--ELRAPELEHGLD--- 1646
Oy 1333 QLTGSGSPKRSQSLISLQOQVYSILOQOALADROHCEV 1363
Db 1647 ----OGLTKTEADQCRHILQVYHSLQVADYAHNDQOLRL 1683

RESULT 8
US-09-452-294-1
Sequence 1, Application US/09452294
Patent No. 6287790
Inventor: Bissell, Sophie
Applicant: Bissell, Sophie
General Information:
Title of Invention: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
TITLE OF INVENTION: THERAPY AND DETECTION OF PROLIFERATIVE AND
FILE REFERENCE: IB-1454- Sequence Submittal
Patent No. 6287790
Current Application Number: US/09/452,294
Prior Filing Date: 1998-11-30
Prior Application Number: 60/110,420
Number of SEQ ID NOS: 1
Software: Patent In Ver. 2.1
SEQ ID NO 1
Type: Protein
Organism: Homo sapiens
US-09-452-294-1

```

```

Query Match 7.3% Score 514; DB 4; Length 2101;
Best Local Similarity 20.1%; Pred. No. 2,9e-20;
Matches 309; Conservative 274; Mismatches 524; Indels 434; Gaps 61;

```

```

Oy 549 RNQIKQNEMLVEEFKR---DEGKIMEENKRIQLEKLSMCE--LEBERKORLITPMBQOLD 603
Db 677 ELQJHSEDOQKATKERYVAQEKQDQLOBOLOALAKESIKATVJSGLEBERK--RAAOALDEQDNC 735
Oy 604 LSAKALISLPAEFENKSLISNELMEKAKLIDVEREYNSLN-----ETRP 651
Db 736 IS-----ELAKE---TRSLV--EQKREKRELEBERAQRQALAILQJENQATREYV 784
Oy 652 LKRELENIKAKLAQHYVPEHNDQKSLR-----KQSGELOKRTITELISNQTLO 701
Db 785 LRREL--AEAMAQHTAESECEQDLYVEAAMRDQCEQSOGEAOYAMFOE---QMLTK 839
Oy 702 KEIEXVCILDKLLITQOYNNLTETKKNVPLVYSEBKKSCHDYVVDLANKLSDVTHKYTEK 761
Db 840 ECKCKAQREQLQEMEKVACISHS--LQISQKNN--LAEHLNAILAQLQVQK 891
Oy 762 KLEMEKL-----LMEX-ASLSKKNVSLRLEV-----IPPR 791
Db 892 EVRAQKADLQSLQEMAKTSKEVARELTVLRKAGBOOETAENELVEPAPADQRPMP 951
Oy 792 HEK-----WMAKSNTEILKQSLSEL-----KKKQEDERTVYSLMS 829
Db 952 LEEQOQFQCSQVQAALQNEKMBQNGHELELRALMELSGOQOQOEDQOQREVALQ 1011
Oy 830 ENN-----DKTKSHQYQVYKHEETITSLTDKTRRLDYK- 870
Db 1012 ERQROADLALKEAARAELEMLQNALNEQVREFTILOPALLAHALTEKQKQELAKNG 1071
Oy 871 -----KKQDINOEFVKINDVELIKRNLNTQNO-----YKAEV 905
Db 1072 LEAQOITELIELAQYKOLNE--QLANKEKHAQSGQASGAGTEPTGALBLARLV 1129
Oy 906 ISL---REBERKSGILPKSKKQVYDQNSAF-----ILAYKSGOERTVILHETIAOK 954
Db 1130 SKLEQOQKQOQOQOQDLSERSLEAEASNAERQSALFETQOQLEKAAQEGLSQALASQ 1189
Oy 955 RELUTIOET-----KLTVAPITISLE---CERKFKATE--K 986
Db 1190 RELAAFRVQVQKHAQEDKMAQVQKQOBAERKNSLISLSEEVSTLNBQVLEKRGSK 1249
Oy 987 ELKEQLOSGQTKYNTSEERAKKQNDKIKKEITLQ-----KOLKKNWYIE 1035
Db 1250 ELKRLVASEESQKLEESQACQROPATVPELOMAALQGRGRASGRAEKQVASE 1309
Oy 1036 NSVETBRALSRKTEELNROLKIDLOLTKTEAKKEKEL--VEENAKN--TESLIAQVLLQ- 1091
Db 1310 NLQOELTSQNEVAEQLQDELQALQNDKPEQFOQALSTLQELNJTQALVSLLEKPIQLOO 1369
Oy 1092 ----KQHVPLQVYELSKISQSTFETLKEELIKQK-----CYEKQOYL 1132
Db 1370 LQAEQQAAREKR--RELEQSKQAQAG--LRAELEAQRELSLPIQVQARQDETA 1423
Oy 1133 TQLR-----QMLEQKNSVPLALEM-----QVKEAPKPVQI 1166
Db 1424 QQLAEKASVAYQSLKAKKAGLLEENKNGLEBRANLQFLEVLOQAREKQVQALAV 1483
Oy 1167 KAS---LEKREPSQNTTEYSKQSELOIQTQKALKTETREVVQDSTYATKSLDQTO 1222
Db 1484 RAAATSTLQVQREAOQSTAELETVATKAEQAK--VKYLEQO--RPQERQKQLAQ 1536
Oy 1223 ISDLNERTKLNMRKYEEVCEVLAHAKKKELSA--KDKKELHPFSTIE----- 1266
Db 1537 VEELSKRLD-----SDASRVOOQKILAVNOAGSGSOQAPROQADLELOAOISOKEO 1591
Oy 1267 -OETKQOEDCKSLITTELOPOTLESANQOE--ADNDKTELLINDVETLOKALNGIS 1322
Db 1592 ALENKILQNEKATKTHYDAKKNQOQOEDLOPOLSLEDOLEKEM--ELRAPELEHGLD--- 1646
Oy 1333 QLTGSGSPKRSQSLISLQOQVYSILOQOALADROHCEV 1363
Db 1647 ----OGLTKTEADQCRHILQVYHSLQVADYAHNDQOLRL 1683

```

```

RESULT 9
US-08-195-487-4
Sequence 4, Application US/08195487
Patent No. 5763403
GENERAL INFORMATION:
APPLICANT: TONKATLY, GARY
APPLICANT: LUDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL HALLMARK CELL TYPE MARKERS OF THE
NUMBER OF INVENTION: 6
NUMBER OF SOURCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZITE: 02.109
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,487
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/901,701
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESO, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MFP-013
TELEPHONE: 617/248-5100
TELEFAX: 617/248-5100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-195-487-4

```

```

Query Match 7.3% Score 512; DB 1; Length 2101;
Best Local Similarity 20.1% Prod. No. 3.76-20;
Matches 309; Conservative 273; Mismatches 525; Indels 434; Gaps 61;

```

```

OY 185 KONTALMAGCGKDAVEY-LIKNGADVTLLDALGHSSTYARIGNDLITLLTKAS 243
DB 215 QMRRLKQADGNSNDELEJELAKKLTETKDA-----QIAHQQRIRLALNKNK 268
OY 244 ENS-----NGRGELMKKGPSLOOR--NLGQMLDEVYTSNMOGEHONQDLEIENDEK 296
DB 269 ASPLTEKELELHDKNESLYNRHLETLKQCCD--RKINQLESEKNDLSFR 326
OY 297 LRRTOEDRRLILDKVNGU-----QULNEEYVVAADLESEK--EKLSULA 340
DB 327 LRPVSHLQADQDLMLNETLEHSHKATQEWELKQKQULESLNAN--LDQKKCLEKNEILL 384
OY 341 AKRQHEESLRTIE-----ALSRFYKPFESH- 367
DB 385 GKLSQLEHLSQLDNPNQEKGRVLGDVLTQLEKQEAATLAANNQDQAEVLELETKRG 444
OY 368 -----LSSGSIFKEDMLKQ-----GQWYMTD 390
DB 445 QDEANLLAERGHPEEKQOJLSLIDGSSQ ISNLSQASKEELDQASQHGAKRLINQVASELT 504
OY 391 SCCTGMYHMGSSRLRPLRLAPQKASVSEMLIKKLEIPAMPTQDSQAKORLQIKAN 450
DB 505 SEUTLTMATIQGDDQELAGLKQAKERQKQALQAF--LDQDSQASQGLRHQVEQ-----LSS 558

```

```

OY 451 ELAHVAVBCALALECEBRYEKEDSDQ-----TKOLEDALQKOVQRMVYESGKVK 499
DB 539 SLKQKQDLKEVYAKQKABRDQADQALQATMEERBASILEKQALQKLEKRAKKE 518
OY 500 QMOTHFALKEMLTSDATQ-----HKLMEELKQULKQMKYK EGSASAPYGL 548
DB 619 ILQDQULQANE--ARQAGTSYVQNGRKAELKSLKRYELQACVETARQDQNGQAVRL 676
OY 549 RQIQIKQEMULVEEFR--DGKLYEPRKQLKTELSPCE--LREKRRQLRLEMEQGLD 603
DB 677 ELURSEDOQATEKRYVADEQDQLEQDQULKESTLYVTKGSLSEER--RAALALDQNC 735
OY 604 LSKALSLSTPKEBENKSLSELNEKANKLIDIVERYSIN-----ETRP 651
DB 736 IS-----ELAAE--TSLV--DQKREKRELEBMRQKGLERALLQDGLAMQAFETV 784
OY 652 LKRELENLAKKLAQHVPRPEHQLSRV-----KQSGEIGKRIITELTKNOTGL 701
DB 785 LRREL--LEMAAQHTASECEQALVEYEAAMQYEDSDQSEAGQAFQCE--QIMTLTK 839
OY 702 KEIEKCDLKKLLTQOVNLTTEKKNVPLVSEBKKSHDYIYDMLKSLSTYHTKTK 761
DB 840 EESKREKQLEQENKQKAGLISHSE--LDQISQNK-----LAEILNANLAQAQOKR 891
OY 762 KIKRELT-----LMKV-ASIKSNVSELETV-----IPPR 791
DB 892 EVRAQKLADQLSTQERKMATSKVRLLETLYVRKKGQOQTASRLVEPAPAKQDPRW 951
OY 792 HEKE-----MVALKSNITELKQULSEL-----NRKGEDEKITYLSMS 829
DB 952 LEBQGFQFSTQALQDNEBQNGHELELRALNLSQGGQDQENQDQENQAVARLQ 1011
OY 830 ENM-----DLKTNSHQVQPVYHEETLSTLSTLDTTHRELDVNYK- 870
DB 1012 EBGQAQDQLAEKRAARELEKRLQNALMDQRPVQFATQALALAHTEKQDEQLAFLNG 1071
OY 871 -----KQCEDINQEFYKIKQENILKLNLENTQNG-----VAAEY 905
DB 1072 LEMAQIKELBELKQTVQKLE--QIAKREKHAQSSQASQENQAGTERTQKALNDAEY 1129
OY 906 ISL-----REIEKMGSLRKSQKQVQNSAR-----LLAKYKSGQERTVTLHEHIAQK 954
DB 1130 SKLQGGQKQKQDQSLERSLEAFRASAPAEBSALETLQGLQERKQELGHSQALSASQ 1189
OY 955 REIDTTOECI-----KIKYAPITSLSE--CERRKPAE--K 986
DB 1190 RELAAKRTVQVDSHAKADQKAVQANGQKQENKKSLSLSESLVILMRYLEKDESK 1249
OY 987 ELKEQDLSQGTQKNTSEENAKKCKQDQNKLIKETLQ-----EDLKQVNYHE 1035
DB 1250 ELKTLVAMSEBSQKLEBSQACQHQVQAVPVLQNALMLGCRKCAQSGEAPKQVASE 1309
OY 1036 NQVETSRALSKRTTELNRQJLQJLYEAKREKREK--VEEANAQ--TSEILAAQTLLO- 1091
DB 1310 NLRBELTQSMARAELEQELKAMQEBEFTQKQMSLTQLEHTSYQALVSELPAHNIQDQ 1169
OY 1092 -----KQHVPLEQVESLAKSLSGTITELKELTKQK-----CYEKEQV 1132
DB 1370 LDQAQAAAEKHH--RELDQSQNAQAG--LPAILLRQRELGELITLPQKQVABORPTA 1423
OY 1167 KAS-----LREKESQNKTEEYSKLOSETQNTQKQALKLETRVINDLSKTKATSDLEL 1222
DB 1484 KQDETRLEQVQNGQSTARELEVPYKATQKAK--VAVLEERQ-----RQDERQKQLQD 1536
OY 1223 ISDNLKELNANMKRYEEVCEVYLAKKKEKLSA--KQEKLLHHSJSE----- 1266
DB 1537 VEELSKKRLAD-----SDQASKVQOQKLAQVQAGQSDQDQAPRQQAQMLNQLQALQSKRQ 1591
OY 1267 --QEIKDQDERCKDSLTTTELQARRIQESAKQIE--ANQNKITELLNDBERLQKANGLS 1322

```

```

Db 1592 AAEHKKQMEKAKTHYDAKKQONQELBQBLSLQLOLEK---ELAKBEMKSHQLO--- 1646
Oy 1323 QLTGYSSSPKRSQSLDLSQOQVHSLQOQLADADQHOEY 1363
Db 1647 ----QGLATKBAQTCRHL7LQOVHSLQOYAAHADQLOLRL 1683

RESULT 10
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TUMOR SPECIES: INTERIOR NUCLEAR MATRIX
; NUMBER OF SPECIMENS: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; PCT DATE: 19930621
; CLASSIFICATION:
; APPLICATION NUMBER: PCT/US93/06160
; CURRENT APPLICATION DATA:
; SOFTWARE: Patentin Release #1.0, Version #1.25
; OPERATING SYSTEM: PC-DOS/MS-DOS
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ. EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; FAX: 617/248-7100
; INFORMATION FOR THE USER:
; SEQUENCE CHARACTERISTICS: 4:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-06160-4

Query Match 7.3%, Score 512, DB 5, Length 2101:
Best Local Similarity 20.1%, Score 20, DB 20, Length 20:
Matches 309; Conservative 273; Mismatches 525; Indels 434; Gaps 61:

```

```

Db 505 SELTATLNTTIOGDOETLQIKQOAKKQOAOLOAT---LOQBPBAQDGLRHQVQ---LSS 558
Oy 451 ELAKRAEKKALALECEKVEDSDQ-----IKQLEBALDVQRYMESEQKY 499
Db 559 SLAKKQOQLKVAEKQATROHQAQOLATMAEPBAELBQALQOLELAKKAKLE 618
Oy 500 OMOTHFILAKHEHLSDAATGN-----HRLMSEKQOLKDKKXKHEGASQVCL 548
Db 619 ILQOQLOVANE---ARBSAQTVTAQARKAELSRVEELQACVTAQROBQEHQOAVEL 676
Oy 549 RHOIKQEMALVEEPR---DECKLMEKKRLQELSCNE---LEREKORLTEMQOGL 603
Db 677 ELQABEQKATYERKRAQOEDQLOQALQKSLAKYTGSLSEER---RAQALDQOQK 735
Oy 604 LSAALSLIPAKPEFNKSLSNGLMEKAKLIDYERESLNL-----ETRP 651
Db 736 IS-----ELKAE---TRSLV---DQKKERKELEBERAQRGLAALLQGLAQOAFBY 784
Oy 652 LKRELEKAKLAQVAPREHDLKSRU-----DQKSGELKRIHTELSTNOGL 701
Db 785 LREPL---NEMAQOATSECEQOVLVYEAAMQOTEDSQOEEQOAMQOF---QMLTK 839
Oy 702 KEIEKQADKMLTQOYNALITTEKKNVPLVSEEMKSHIDYQLOLXKLSQYHAKYK 761
Db 840 BECEKAOELQAEKAKAGATSHSE---LOLSNOOK-----LAEHIANIARALQOYK 891
Oy 762 KLEMERU-----LKEN---ASLSKNVSLTFV-----TPPR 791
Db 892 EVRAQALADSLTQERMAKATSEVAVRLTLVRKAGQOETASRELVKAPARADQBP 951
Oy 792 HEKE-----HMAKSNITELKXQSL-----KKGQDQKATYSLMS 829
Db 952 LERQOQPCSTOALQAMEAREADQWQELERLBAALMESQOQOQOERQOEREVATLO 1011
Oy 830 ENN-----DKTNSHOVVPKTHEBITALSTLQTKHRLVDYK- 870
Db 1012 ERGRQADALAKRAKRELBRLOANLQEDQVRETRLOALALALTEBCKQOELAKKLG 1071
Oy 871 -----KKCEDINOEPVTKDENEELIKRLNLTQNO-----VKRY 905
Db 1072 LEAADIKLELEIQTVOYKLE---QAKKEKHEBASGQASQEAQSTPGRPLBALRBY 1129
Oy 906 ISL---REHEKYSGLKRSKQKVDQNSAE-----ILAKYKSOEIVTLHEEIAQK 954
Db 1130 SKLEBOQKQOQOQDQSLERSLQEMAKSHQDSALQETQOGLQEMQOELHQSALASQ 1189
Oy 955 RELDTIOEC-----KLYXAPITSLF-----CEKFAKTE---K 986
Db 1190 RELAFRTKYQDHSKAEDEKMAQVAVRQOAPRNSLSSILEEVSILNRQVLEKGGSK 1249
Oy 987 ELKQOLSOQTOXYTSEBBAKQKQENDIKKELTLO-----KOLKQKVHIE 1035
Db 1290 ELKRLVAKESQKLESCACQROQXATYKAKKEKLE---VEKANO---TEELIAQLOL 1091
Oy 1036 NSRETEPAKSPYTELBNRLQLOLQXATYKAKKEKLE---VEKANO---TEELIAQLOL 1091
Db 1310 NLBOELTSQAEPALEIQOELAKQOKEFPQOBALSTLOLBHSTQALVELLPKHLQO 1369
Oy 1092 -----KQHVPLEOVESLAKRSQSTLEPLKELBKTQR-----CYEKQOQV 1132
Db 1370 LQOQQAALKRNI---RELBDQKQAGG---LRRELAPQRELAGELPLRQKVAQBDQTF 1423
Oy 1133 TQLR-----QMLEQKNSVPLAEL-----QVKAQEKQVCI 1166
Db 1424 QOLRAEKASVAYBOLSMLKAKHGLAELBNQGLQENALMLQROPLEVEDLQAKKXVQGLA 1483
Oy 1167 KAS---LAEKBEESQNTTEVSKQSELOEYQNTQKALKLETEVEVDLSYKAKRSQLE 1222
Db 1484 RADAETRLAEVQRAQASTAELEVTAKYKAEAK---VKYLEEQO---RFOERQOKLTAQ 1536
Oy 1223 ISDNLATNLANKTEYCEVYLIKAKKELSA---KQKELHLHFSTB----- 1266

```

Db 1537 VEELSKRLAD-----SDASKYQOQKLAVALAGGSSODEACROFQOALNELQALISQKRO 1591  
 Oy 1267 --OETIKQFPCCKSLTITTELORRLOSANOIE--AKONKITELANVERTELKQALNGLS 1322  
 Db 1592 ABEHYKLEKAKATYHAKKNOONELQDQLASLDELQLEKMK--ELAAERLGHLELO--- 1646  
 Oy 1323 QLTGSGSPKRSQSLDLSIOQOVSRLQOQLADARQHOEV 1363  
 Db 1647 ---QAGLTKREAEQTCRHILTAQVRSLEAVALAQAQDLROL 1683  
 RESULT 11  
 US-08-592-126-148  
 ; Sequence 148, Application US/08592126  
 ; Patent No. 5821091  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gregory Doljanov  
 ; TITLE OF INVENTION: Transcribers Encoding Immunomodulatory  
 ; TITLE OF INVENTION: Polypeptides  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDING ADDRESS: 51  
 ; ADDRESSEE: Dehlinger & Associates  
 ; STREET: 350 Cambridge Avenue, Suite 250  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94306  
 ; COMPUTER READABLE FORM:  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/592,126  
 ; FILING DATE: 4/25  
 ; PRIORITY DATE: 4/25  
 ; INVENTOR INFORMATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sholtz, Charles K.  
 ; REGISTRATION NUMBER: 38,615  
 ; REFERENCE/DOCKET NUMBER: 4600-0111  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 324-0880  
 ; TELEFAX: (415) 324-0960  
 ; INFORMATION FOR SEQ ID NO. 148:  
 ; SOURCE: GenBank  
 ; LENGTH: 1312 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEetical: NO  
 ; ORIGINAL SOURCE:  
 ; INDIVIDUAL ISOLATE: Rad50, pro-translation of SFO ID NO.54  
 ; US-08-592-126-148  
 Query Match 6.84; Score 477; DB 2; Length 1312;  
 Best Local Similarity 22.1%; Pred. No. 1.7e-18;  
 Matches 314; Conservative 236; Mismatches 515; Indels 356; Gaps 70;  
 Oy 108 LOYNGPTEH-VDLQGRALHDAAM---DQSSITGLCDMGASVNAQDQGTLYLQNO 163  
 Db 49 LKLTCTDPPPTGKNTFYNDPVVAQGTIVRAQRLR-----QFQDNGCF---LIAYQ 97  
 Oy 164 MCPPTICQLIDRGADINSRKQNTALMGCEYGCNAVAVLIMGADVTLLDAIGH-D 222  
 Db 98 --RSWVC-----TQKSKRT-----EKTLEGVITRTGHEVSLSSKCAED 137  
 Oy 223 SSYVATIGQDLIL-TLLKTASNSNGRELMKAGSILQOQNLISQMLDEVNTKS----- 275  
 Db 138 REMISSLQASVKNLVITGHQDSN-----W---PLSEGLAKQXDEIDPSATRIYAL 189  
 Oy 276 -----NDRPHQNTQDLEIENEDLKERLKLQGOQRLILDVYNGVQQLQNLNEEVAADIE 329

Db 190 ETLRQVROGQKAYKEVQMEKLYKQYKEACE-----HQDITSKEQALTSRKREIYKSE 245  
 Oy 330 SEKEKIKSLIAKKEQHEESI-----RTIALKREPYEESDILSGSHREEMILIAQ 384  
 Db 246 NEIDPLKRLR--KELEHNSIKIKLQNDLALDSRKQKQKEND---SELEKKEVPOG 299  
 Oy 385 QMTVRODCTSTGMPVHMQSNMLRLELALPDQASVSENLLEKLELMAATITCDQSNKD 444  
 Db 300 -----TQDQLN-----DLVHQRYVREKE--RLVQDCHRELKLNKES 336  
 Oy 445 RLKIQBELAKVAECBCKALACECEVENE-----SDQRIQ 479  
 Db 337 RL-LINQKSELIVRQGLQADRHQDRIKANDSLQSLATQLELDGCEPHGFSEHQIKN 395  
 Oy 480 LEQALDQVQRMYSEBCKQMOQHEFLAKELHTS-----DAGTCHRLKELADQIK 532  
 Db 396 FHAKVREHQ-----RQKNAQGLKMPAKFETLLQKQNDIDELROKTLGAGII--ELASEIL 450  
 Oy 533 DMK-----VKY-----EGASVGLNQL--IKQREML-----VEEKKDEKQKME 572  
 Db 451 SKRONELKNVYELQOLESSDRILELOQLKIKARELSKAKNSVTLKMEVSLQME 510  
 Oy 573 NKRIQRELSMCELEERERGRKL--TEMEQQLDLSAKIALSPAKFEMKSLLSNEN 629  
 Db 511 KADIDTILKLDQEDQMDQHMTTRQMDHRTKQDAK-----DQQLRKISHSBEL 564  
 Oy 630 E-----KAKKLIQREYERESLNETRLKRELEMLAKLQGHKPEEH--EQLSRLQ 681  
 Db 565 SILQYFPKKQLBMLHSKSEIKNOTRD--RLAKLNKELASPSQKKNHINELKREBO 621  
 Oy 682 KSGELKRL-----TELTKNOTLOKEIEVQCLDKLL-----TQGVNLTTEKMH-- 727  
 Db 622 LSTEDRLQDYGSODPESQDRLKKEETLSKQAMLAGATVYSPQITQDTPDQNSC 681  
 Oy 728 --PLAKVSEMKSHQYVQDNLKXLSDTHTKTERKLEKELKLEMASSLSKVSRLQYF 786  
 Db 682 PQCQRFQVTELOEVI--SDQSLK-----RLADPKIKLSTESLKK-- 722  
 Oy 787 IPEPRHEKEMAL--KSNITELK-KQLSELNKKGQDEQKITYLSMSNNQDLKTKMSHQY 842  
 Db 723 --EKROEMQLQVPMQSTIDLEKEVLEPKELKQVNVNDIQRL--KADIEQETILG 776  
 Oy 843 VAVVTEHEELTKALS--STQDKTNEELVYKPKCE-----DINQEFKVIDENELIK 891  
 Db 777 TITPEESAKVCLQVYITERROEMKEVEERKLAQAAKILQDILQRTVQVQVQDQKQ 836  
 Oy 892 RLNLQNTQVQVAYISLRHEKMSGLKRSKKNVQDSAEI--LAKYKKSQETVTLHE 948  
 Db 837 KHLQVYSKTELRKLQDQDQDQLOHLSSTNLELSEKQLQISTHLDQHQDQLEQVYELST 896  
 Oy 949 EPLAQQRRELDIOECIKATYAPISLEECRPKFKATKELKQLSQOQYKYNSEPAEK 1008  
 Db 897 EYQSLVREIKQAK--OVSPELTTEKFOOEKEEL-----INKNKSNIAQ-- 941  
 Oy 1009 KQOENKIKLELITLQDKLD--KNVH-----TENSVTERALSRTTELEQLQDLQALQ 1061  
 Db 942 -----DKL-----NDIKREKVNHNKGMDIENITDQKQDQKQKE--TELKAYIQ 986  
 Oy 1062 YTEKKEKREKYEENAKQVSEILQALQLOQVHLDQVLE--LAKSLSQRTIELKELK 1119  
 Db 987 LSECEKREKINED-----WRLMMDIDITQKIOEPMHQLQNL--TLKRNNEELK 1032  
 Oy 1120 -----TKQCYEKEDQYITOLR--QHLN-----QKNSVPLAEHLQVKEAREKVI 1166  
 Db 1033 EYEEERQHLKQEQMOVLQMSHQDLEENIDNIRKNNHILR--QAGIEELIHF 1088  
 Oy 1167 KASLREK-----RESQNK-----TEPYK-LOSEIQTQKALKLETFEEDVLSYKA 1214  
 Db 1089 KKEILREQFRADEKXREPMHITVHETELVKNQDILTYITLQDAIMFHSKMEELNK-- 1146  
 Oy 1215 TQSDLETOISDL--NEKLANIKREYECVLELAKKRSLSARDKEILHFSI-----RQEI 1269  
 Db 1147 -----TRDLWRSTYRGQDIEVITERSDA-----DENVASDKRRNWTIVVMLKQDTA 1194

```

Oy 1270 KDOORCKDSITTELORIOESAK-----OLEAKONKTTTEL-LNDBERLKOALNSLO 1323
Db 1195 LMKRGCRSGOKVLAISITLRLAELFCJLGLDEPTLMDRENIIESIALH----- 1249
Oy 1334 LYTGSSPSKRSQSLIBSLDOQVRSIQDGLADNRQOEVT 1364
Db 1250 -----VEITRSBQ-RNQOLLYTHDEDEYVL 1277

RESULT 12
US-08-687-080-51
Sequence 51: Application US/08687080
Accession No. 568170
GENERAL INFORMATION:
APPLICANT: Gregory Dolanov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CHROMOSOME LOCATION DATA: 08/687,080
APPLICATION NUMBER: 08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Shultz, Charles R.
FIRM: Shultz, Charles R.
ADDRESS: 38,615
TELEPHONE: (415) 324-0880
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
SOURCE: amino acid
STRATEGY: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 6.7% Score 474. DB 3. Length 1112.
Best Local Similarity 22.0% Prod. No. 2,5e-16.
Matches 313: Conservative 237: Mismatches 515: Indels 356: Gaps 70:

```

```

Db 138 REMISLIGVSKAVLNIVTECHQEDSN-----W-----PLSGKALAKOPDEIFSAFRIYKAL 1389
Oy 276 -----NOREHON-IOLELEIEMEDKERRLRITQOORLILDRVNGILOALNEEYVADLE 1329
Db 190 ETLRGVROQGVQKVEYQVOMEIKLQVYEEKACE-----IIDQITSNKQDLSEKTEYKSTE 245
Oy 330 SENEKSLSLAAKEKQHEESD-----RTIETALKSNFPEESDILGSSHREKREMDLNG 384
Db 246 NEDLPKANRL--KEIEHNSIKIMKLNDIEALADSRKKQEMEDN-----SELEKKEKVEYFG 299
Oy 385 OMWYDSQCTSTGAPVHMOSSMLRPTELALPNAQSYSENEILKELEDAARTFCDANRD 444
Db 300 -----TDQLN-----DLTHMQRVYNEKE--RKLVDCHRELRKLNKXS 316
Oy 445 RLALQELHAIKACSKALALECERKED-----SDDQIO 479
Db 337 RL-LNDRSESLVDSGLDLOADNRQHEIHARDLSLOSLATOLEDFERGSPFSEQIK 395
Oy 480 LEOALQDVQKRYTESBQYKQOTHELAKELTNS-----DAATGNHRLMEELDQDLK 532
Db 396 FHLVLRNQ--EGKRNQALQMDNREKLETLQKQIDETIRKTKTGILH--ELKSEL 450
Oy 533 DMK-----VKY-----EGSAEVLKRNQ-1KONEM-----VEEPRQDGLME 572
Db 451 SKKQNELKNVKEYEQLDEGSDRLLELDQELIKKEELSKKESKNVETLKMEVLSLQNE 510
Oy 573 NKRLQELSMCELRERKQKRL-----TEWEGQDLQSLAKLALSTPAEFENKSLSLNTEL 629
Db 511 KADLRLRLKLDQMDQJLNNHTTTTQLEMLKQKQKQK-----DEQIRIKSHSDLT 564
Oy 630 E-----KAKTLIVREKERSLNETPLKRLLENLKKAKLAQHYPREN--DOLKSRLEQ 681
Db 565 SLGDFYFNKKOLEDMILSKSEKINETQD--RLAKLNKEASSQNKNNHINNELKKEQ 621
Oy 682 KSGELGRN-----TELTSKNQTLQKEIEKYQDNLNLT-----TOOVNMLTEPKN-- 727
Db 622 LSTEDLQDVPDQSGOPESDLDLLEBELIKESKORAMQANVAYVSYOTDQLENGSCC 681
Oy 728 -VPLVEEEMKSHVDYIDLANKLSVYHNKYTEKLEKELMKNASLSKNVSLRLETV 786
Db 682 PVQGVQTEAELOEVI-SQLQSL-----RLAPKXLRTESELKK----- 722
Oy 787 IPERHEKEMAL--KSNITELK-KQLESELNKKCEQOEKYSJMSJENMLKPKYSHOY 842
Db 723 -----EKRRQMDGLVPLRQSLIDLKKEIPELRKNLQNVNDIORL--KNIDQEGTLIG 776
Oy 843 VPKTHKEIKTALS--SLQKTNELLYDKKCE-----DINQEVKIDENELTK 891
Db 777 TTPREESKALCNDVITMERQMDLNDVDEK-IAQAAKLDGLGIDLDVQVQNDQKQ 836
Oy 892 RINLETONOVAEYISLREHEEKMSGSLKMKKQVQNSAEI--LARYKSOSEIYTLHE 948
Db 837 HKLDVSSITELNKLIDQDQOIOHLKSTNELSEKLS1STNLRQOOLEQVYELIST 896
Oy 949 ELAQAQLEDTIORCIKAPLISLSECEKFRATEKLEQSLQOOTKQYNTSEBAKK 1008
Db 897 EVQSLYTEIKDAKE-----QVPSLETLTKQKREBL-----INKNNSKKNQ- 941
Oy 1009 CQKQNDLKEILLOKDLQD--KRVY-----TENSYTEFRLASKEITELRLQDLQK 1061
Db 942 -----DL-----NDIKEVKNHIGHMQDENLYIDQDKQVYKQKE--TELKLVND 986
Oy 1062 YTEAKKKEKLVESNAQOTSEILAQTLLOHVRPLEVES--LKNLSGTTETLEKELK 1119
Db 987 LSECKMKKIKND-----KRLMKDQITQKLODENHLDQNL--TLKKNREBLK 1032
Oy 1120 -----TORCERKEQOQVYTLQ--QKLEN-----QKSSVYLAEMHLYQKRAKREKII 1166
Db 1033 EYEEERQHLKEMQGOVLOLHMSSEKLEENIINKRNHNLAL--QKQGESEIIEH 1088
Oy 1167 KASLAREK--EESQNK-----TEEYSK-LOSETQNTQKALKTLETRVVDLSYXA 1214
Db 1089 KKEIRLEQFQBAKEKYKREMYKMTTELNVNLTLYTKTLQDLAKMFKSHMEIINTKI-- 1146

```

OY 1215 TKSDELTOISDL-NEKLANLNRKRYEVECEVLIHAKRELSANDEKELLHST-----BOEI 1269  
 DB 1147 -----INOLNSTTROODLEIETISDA-----DENVSASDKNNNTNRVYMLKODPA 1194  
 OY 1270 KDOOECKOSJTTITTELOER-IOESAK-----OIEAKONKTEL-LADVEPLKOLNGLSLO 1323  
 DB 1195 LDMGRCASOKVYLASIRILALEFPCNLGIGLIDETTNLDRENTESLAAHL----- 1249  
 OY 1324 LTVSGSGSPSRKOSQIDSLQOQVHSLQOOLADADROHQEVI 1364  
 DB 1250 -----VETIKSRSDO-KNFOLLYTHIDDEVELL 1277  
 RESULT 13  
 US-09-104-324B-4  
 ? Sequence 4, Application US/09104324B  
 ? Patent No. 6232460  
 ? GENERAL INFORMATION:  
 ? APPLICANT: T recti, Ozlem; Sahin, Ugur; Pfeunderschuh, Michael  
 ? TITLE OF INVENTION: Methods for detecting circulating cancers,  
 ? TITLE OF INVENTION: Methods for identifying pathogenic markers in a sample of  
 ? NUMBER OF SEQUENCES: 4  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Fulbright & Jaworski LLP  
 ? STREET: 666 Fifth Avenue  
 ? CITY: New York City  
 ? STATE: New York  
 ? ZIP: 10103  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
 ? COMPUTER: IBM  
 ? OPERATING SYSTEM: PC-DOS  
 ? SOFTWARE: wordperfect  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US-09/104,324B  
 ? FILING DATE: 25 June 1998  
 ? CLASSIFICATION: 435  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 08/892,702  
 ? FILING DATE: 15-July-1997  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Hanson, No. 6232460man D.  
 ? REGISTRATION NUMBER: 00,946  
 ? REGISTRATION DATE: 01-01-94  
 ? TELEPHONE: (212) 318-3000  
 ? TELEFAX: (212) 753-5958  
 ? INFORMATION FOR SEQ ID NO: 4:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 976 amino acids  
 ? TYPE: amino acid  
 ? TOPOLOGY: linear  
 US-09-104-324B-4  
 Query Match 6.5%; Score 454.5; DB 4; Length 976;  
 Best Local similarity 22.3%; Pval. No. 2.1e-17; Index 319; Caps 51;  
 Matches 256; Conservative 199; Mismatches 376; Indels 319; Caps 51;  
 OY 219 LGHDSYVARIG-----DWLDLTLKTRASBNSKNGELMKKPSLQORLQSLQMLQOMLQNTK 274  
 DB 28 LGDSYTFKSPKCTEDLE-----PPAKTNLSKNGEMISDPAQKVFPLVLEQV-N 82  
 OY 275 SNORHONIQDLEIEN-EDLKERLKLQOEOHLLIKKNGLOLQINEVYVADULESEKE 333  
 DB 83 SDOHVOGCLKSDLENSGSLRVSFKLYE-----AEKIKKAVY 121  
 OY 334 KKSLLAEKQHEESLRTTBALSKREKYPFESDGLSSHPKREDMLLKQSQMTMDSOC 393  
 DB 122 STEALNQKESLQDENMKKILBQAKQIQELQ-----FGNEKVSULV----- 162

OY 394 TSGCPVPHQKSRKRLPRLTALPBNQATSENPILLKELAMPTGDSASQKLOKUNELA 453  
 DB 163 -----EEIQENKNDLIRENNATRHLCNLKKE-----TCA 191  
 OY 454 HVAACSTALALECEBYKE-----DSDEQIKOLEJDLAKOVOKRYESSGCVYKQOTIFLAK 510  
 DB 192 RSAETKATYERREBTROVYDLMNNTKNTIYAFDLOAKENS-----RLBNH-TKLE 245  
 OY 511 HLTSDAQNGHMLBKLKQOLKMYKYIYBSASAKYLRQIKOMKYEKPRKDSGKLA 570  
 DB 246 -----DYEKIOL-LQDEYKKEIENKQKVSLLIQTIEKNNKMDLTLELE-SDDKVN 299  
 OY 571 EENKRLQKELSMCLBEEKRORKLTEMQGLKALSLIPAEKFEKMSLLSEMLE 630  
 DB 300 EERTKLOSENKOSIERO-----HHLTK-----ELDEIYSLQNSVSTOK-----ALEBQ 347  
 OY 631 KAKLIDERYEYSLNLTREPLKRELNLKALQKRYPREPOULKSRLEKQSOEGLARI 690  
 DB 348 ATITCOLTEKFEQMESEN-----KANAHSYVTFEFTVCSLEEL----- 390  
 OY 691 TELTSNQTLOKEIEKYCLDKMLTQOVNLTLYEMKNVLYSEEMK-----SHDYIVDL 747  
 DB 391 -LRTDQORLEKNEDLO-----KLITLMBLOKSESL-----EEMKRLTKNNKEVELLE 436  
 OY 748 NKKLSDTYHTYERKLDEMKLMEKASLSANVSLFETVTPPERKREKMMAL-----KSI 803  
 DB 437 KKVGLS-----BKETLLKMKOPEKILAEELKOT-----EDLIGILOAREKV 478  
 OY 804 TELKQSLSELNKKGDEOKETYSISENNDLKTYSHQVYVYTHEELK-TALSTLQK- 861  
 DB 479 HDLEIQLAIT-----TSDQYVS-----KEVKOLTELE-----NEKMLTETLSHONKL 523  
 OY 862 --TNREL-----VDKKGCEIDINQEFYKIKQENELK- NLEPNQVAKYIISLR 909  
 DB 524 SLEKRLQSTSDMTLELNQEDINN-----NKQDEMLKQINLQETETOLANE----- 575  
 OY 910 BHEKKSQSLKSKMKRVODNSAETILAVYKQSEIYTLHEBILAKQHELDITQECIKLYA 969  
 DB 576 -----LEVYHEELKQKD-----EYCKLIDKSEBKNLQROVEN-----KNT- 614  
 OY 970 PLSLECEKRFKATENE-----LKEOLSOQOTKYTSEBAKQKQONDKLAKKEITLQND 1026  
 DB 615 -----LELDQENKALKKGTBESQNYEIVKMLELLELSAKQKFGELTD-----TYORE 667  
 OY 1027 LKQDNVHTSEYSEFTRALSRTTELNRDLKOLLQKYTEAK-----KEKELVENAKQSEI 1083  
 DB 668 IEDKLIKSEN-----LLEBYEKAVADENVALKEDEIKDROKR 706  
 OY 1084 LAAOTLLQKQVHLYBQVSLKKSISTETLKEBELKQKQRYEKEQOTYQOLPOMLEK 1143  
 DB 707 IENVALMEKH-----KHQYDMLIEERDESL-----GLYKSKQDSQSRASAEIDL 753  
 OY 1144 NSVPLAHLQVFEAFKEKYGILKASLEKEEESQONSTEYSKQSEIOTYQOLAKLET 1203  
 DB 754 SNL-----KAEELSVYKOLELE-----REKBEKLRKAEKENTATLEKQKQOTFLP 803  
 OY 1204 REYVDLSYKATKSDLETOISDLNEKLANLNRKRYEVECEVLIHAKRELSANDEKELLH 1263  
 DB 804 PEITWMLKQSKAVPSO-----TYSKFTSVSDIGI-----SNDKQDLWT 841  
 OY 1264 STEDEKQKQOGRDQSLT-TITTELOER-----IOESAKO-----IAKONKYTELL 1308  
 DB 842 SAKRTYSLPLPKAYIVYTPPKILOQRNLNIPRESKKRRMAFEFINDSSETDILL 901  
 OY 1309 NDV-----ERLK 1315  
 DB 902 SMWSEEBELK 911  
 RESULT 14  
 US-09-056-200-94  
 ? Sequence 94, Application US/08056200  
 ? Patent No. 5616500



```

GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
INVENTOR: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Method of using protein and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
INVENTOR NUMBER: US/06/056,200
FILING DATE: 06-08-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
FAX: (714) 760-0404
INFORMATION FOR SEQ. ID NO. 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 1898 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-06-056-200-94

```

```

Query Match 6.3% Score 446.5 DB 1: Length 1898;
Best Local Similarity 20.5% Pred No. 126-16;
Matches 298; Conservative 267; Mismatches 583; Indels 303; Gaps 55;

```

```

OY 141 CDGASVAKNDYGRTPVLTATQMCNP-----TTCOLLDGADINSKDKORFALML- 193
DB 23 CD-QALTRKNDLNLREBGAIVLRHRDPVTLDELDR--DSNGRVDFNEFLTF 79
OY 194 -----GEEGCKDAVFLKRGKADVTLLALGDSSTARIQDNIDTLTLTKTASNSNG 249
DB 80 KVAQACVY-----ALGQATGDEKRAKQCKSLDQRTED- 118
OY 250 RELAKGKGSLOQNLNLSQMLDEVNFKNSDRHQNIQDLE--IENEOLKERLRTQIQORI 306
DB 119 -----QRRPRDROLREERQGRNRKORBOBEELAECEGSEGOERLSDQNRQ 167
OY 307 LLDVYGLQQLMEYVADLSEPKKSLLA--AKGQHESSRTITLAKSFR- 361
DB 168 RDEBELMRQ--EMQEREHRAEEDQSCGHELEEFPRDQLRRLELRLKRGRE 224
OY 325 -----YFSDHLAGSGSHFKEDMLLKQOCMYTDSQSTCMVH- 405
DB 262 EKGQQRREDDRVFEE--BEKEMKRTVLAKEEKSLQEE- 405
OY 406 SMLPPE-----LALPNQASTENELIKELDAMRTFDSAKODRLK 448
DB 276 DQLKLELKEQELREERQEQOORLRRQRLRRQGEERRODERRRQOERRRQOER 335
OY 449 QNELAKRVASCAKALALCECRVYEDSDQIQLELDALDVQKRVYEDSKYQKQMTFFLAL 508
DB 336 EQLLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 392

```

```

OY 509 KSHLTSDAQNHRIMBELKQDLKQKVVYEGASAVKLR--NQIKQMLVVEFKRDE 566
DB 393 EQLLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 452
OY 567 GLHEENKRLQKELSMCELR-----EKGRRL--TEHQQLDLSANL 610
DB 453 HQDNHDEERQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 512
OY 611 SIPAKFBNKSLSLNLEKAKKLLDVERVYR--SLMSTRPLKRLN- 662
DB 513 EERRRDEERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 572
OY 663 LQGVNPREHNO-LKSLQKSGELKRIETLSNQTQAEIKVCLQNLKLTQOVNVL 721
DB 573 EKKRLREPRQRLKREDEERKQDLR-DEERKQDLRQDE-RLQDLKREPRV 629
OY 722 -TTEMKNPLKVE--EMKSHDVIVD-----LNMKLSVYTKTEKELKEMKLNMS 774
DB 630 EDERRDEERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 686
OY 775 LSKVYRLTYVTPREHNEK- 824
DB 687 LQRLKREH-----EERRRQDLAEERQANREKISLTKNQMLLSANR--OSKV 738
OY 825 YSLMSENDLKTMSHOYPAVYKTHEITKALSTSLDTNRELDVYKCKCDINOF- 881
DB 739 -----LLEAPQAGRAAPQDEERKRRSEIQLQDEERAHQOQDEERQRPDMQW 789
OY 882 KIDNEVELKRLN-----ENTQVYKAEYLSLRH--EKMSGLRKSKRYQDN 930
DB 790 QREKSEKRGQRLSARPLRQNEHQRLQAEERKQDLRREDEERKQDLRREDEER 949
OY 931 ELAKKRSQEEIYVLEHIAQKRELDTQ- 979
DB 850 QFL-----EERQDLRREERQAOQDEEDGQDQDEERQORRQDKWM-OLEE-ER 899
OY 980 KFK-----ATEKELK--EQLSQOTQKYNSEBAKCKQDNKRLKELLTQAND- 1026
DB 900 KRRHRTIYAKPALQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 959
OY 1027 -----LQKNVHLNENYETFRALSKTEELRQDLQDLQYKAKKREKYLEPENAKT 1080
DB 960 LREDEERKRRREERQYKRRKQDLQCKEQL--LGEERPRRQREKRYREDEELQ 1016
OY 1081 SETLANQTLQKQNPRLQEVQSLKSLSGTITLKEELTKQCYKRRQ- 1130
DB 1017 EQLLRREERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 1076
OY 1131 -----TTLQQLQMLQNO--KNSQVLAELQVKEAFPEVQIYASLPRK--ESQ 1178
DB 1077 EQLLGEERTRRQQLERQYRKEELQDOEERLLEPRER- 1132
OY 1179 NKTVEYSKLOSTQONTQALKLLETREVDSYKVA- 1230
DB 1133 LQDEERQDLRREERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 1192
OY 1231 ANLNKRYEEVCEVLAKKKEELAND--BKILHSLHDEIYD-QQER- 1281
DB 1193 VRQNVVYQKRGNEQFRQLEBSQVYDQSOQDLQHLLEQOERDRQERRRHQANRRP 1252
OY 1282 TITELQRRIOESANQOTLEAKDNKITLELNVYELQALNLSQUTYGGSSPSKRSQDLS 1341
DB 1253 EQLLGEERKQDLRREDEERKQDLRREDEERKQDLRREDEERKQDLRREDEER 1305
OY 1342 LQOQVRSILOQALQADQROHQVQALTYTHLSLAQGH-----ADHQV 1384
DB 1306 REQQLLQDLR- 1384
OY 1385 ALLQITQNRQ 1395
DB 1356 LRLREPRQNLQ 1366

```

RESULT 15  
US-08-800-644-94  
Sequence 38 Application US/08800044  
Accession 5958752  
GENERAL INFORMATION:  
APPLICANT: Steinerl, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Hydrolysis and Transglutaminase-3 and  
NUMBER OF SEQUENCES: 117  
METHOD OF USING SAME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release 11.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
CLASSIFICATION: 424  
PRIORITY: 4 FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/056,200  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Gedick, Michael F.  
REGISTRATION NUMBER: 36,799  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1898 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-08-800-644-94

Query Match 6.3%; Score 446.5; DB 2; Length 1898;  
Best Local Similarity 20.5%; Pred. No. 1,2e-16;  
Matches 288; Conservative 267; Mismatches 583; Indels 303; Gaps 55;

141 CDGASVYAKNDGSGTPVLAATONCR-----TICGLIDRDAVTSRDKONTATL 193  
DB 23 CD-GAALTKKDLKLERFGVALRRDRPTVDLLEIDLR--DSNGVDPMFLLEIF 79  
DB 194 -----CCEYGCKDAVEVLKNGADVTLLDALGHDSSVYARGDLDITLTITLASENSKNG 249  
DB 80 KVAQACTY-----ALQATGIDENRRARDCKESLLQDRTEED----- 118  
DB 230 RELMKKGSFLQDRMLSDMLDEVTASNGRQNIQDLE-----TENDLKLAKIOGRI 306  
DB 119 -----QRFRRPRQLREPRQRNRKQKQDRELALAGEGSEKQERLQDRQ 167  
DB 307 LALKVNGIQQLNEEVAVADLSEKSKELSLLA--AKRKQESLATTLEALKSRK-- 361  
DB 168 RROELIMRQRO---EQMERERARAEEDLOSCGHTEPRFEEDLARRELLLEKRRGE 224  
DB 362 -----YFESDILGSGSHFRKEDMLLKQOGWYDSCISTGAPVH-----MQR 405  
DB 225 EKQOQRREDRQVDFE---EKKMKKEVYLNKEBEKIDEE-----POKQELDEEE 275  
DB 406 SMARPLE-----LALPQNASYSENEILKKELFAMRTFCDASAKORLKL 448

DB 276 EQRLKLEQRLRRQREERQOORLRRQDLRRQREERQREERQREERQREER 335  
DB 449 QNELAKVYAECKALALECEKVEEDSDQIKOLELALQDVQRYVESGKVMQHTPLAL 508  
DB 336 EQRLREERQREERQDLRRQREERQDLRRQREERQDLRRQREERQO---LRR 392  
DB 509 KEHLSDAATGINKHLELQKQIKMKYKKEGASVCKLA--NOKDMPLVEKKEDE 566  
DB 399 EQRLREERQDLRRQDLRRQDLRRQDLRRQDLRRQDLRRQDLRRQDLRRQDEER 432  
DB 567 GKLYEMKRILOEILSMCELE-----EKSGKLT--TEYEQDLNDLSKTLAL 610  
DB 453 HEQKREERQREERQRLKREERQRLKREERTERREERQDLRRQDLRRQDEENHMLIL 512  
DB 611 STAPKPFNNKSLSLNLEAKKILIDVEEYER---SLNTEPTKREPLN-----LYAK 662  
DB 513 EERREERQREERQDLRRQREERQDLRRQREERQDLRRQDLRRQDLRRQDLRRQ 572  
DB 663 LAQHVPEPEIBO-LKSLRQKSGELQKHTLTSKQNTLOKEITLVCLNDKLITQOYNNL 721  
DB 573 EKKALDEERREERQDLRRQREERQDLRRK--EERNRQDLRRQDE--RLQDLKLKEEVL 639  
DB 722 -TTEKNNPLKYSF--EKKSIDYDYO-----LKKLSQVHYHTEKLEMKLMMNS 774  
DB 630 EDERREERQRLKREPRERREHLLKSEDERREHQLRRQREERQRLKRE---EER 666  
DB 775 LSKNVSRLTEVTPRPNHEKEM-----ALKSNITELKQL--SELAKSGEDEXT 824  
DB 687 LEORLKRH-----EERREDELAEEBOQARERIKSLKPMQMLSEADAR---QSKV 738  
DB 825 YSLKSNNDLAKTMSHOYPAKTHETITLSTPLDTLRADLVYKKKCDINQEPV--- 881  
DB 739 -----LLEPQAGRAAPDOEERKRRSEELQWEEERAHQOODEERQDRPTQM 789  
DB 882 KIKDENELIKRNI-----ENTQNOVAYEYLSLREH---EKKMSGLRKSKVQDN 930  
DB 790 QNEKSESGRORLSRPLRDRQRLAEERQDEORLPEERKEORORQRRKREKL 849  
DB 931 ETLAKTKSGDEIVTLMEIAQKRELTIOF-----CIKLKVPYIISLPCR 979  
DB 850 QFL-----EEBQLRRERAGQQLDEEBSGLQDEBQRRRQGEORQKWM---QLER-EK 899  
DB 980 KFK-----ATEKELK--EQLSGOOTQVYVSEBKAKGQENDKLKEITLQND--- 1026  
DB 900 KRRHNTLVAKPALQDLRRQDLQDEEBELQREERKRRRQDEQVQREEDQLQDEEQ 959  
DB 1027 -----LQKKNVHNSYTERALSKTVEELRNQDLQKLTAKKEXKELVFNKQT 1080  
DB 960 LAREERKRRQREYRKDKLQKQDEO-----LGEERKRRQREKRYAYREEDLOE 1016  
DB 1081 SEITLAQNTLOKQHPLEQYESLKSLSGTLETLKELTKQRYEKQ--- 1130  
DB 1017 ESQLREERKRRQREHMYKAKDQDEOEDQLLRREERKRRRQREYQREERLOE 1076  
DB 1131 -----TYTOLRQMLEND--KNSSVPLAEVLQYKAEKVCILIKSLKEF--EEO 1178  
DB 1077 ESQLGGERETRRQDLRRQRYKKEBQLQDEBQDLLEPRK---RQKREHQRREEE 1132  
DB 1179 KTEYESKQSELONTQOLKLTLETWVNDLSUYKA-----TISDLEITQISLMEKL 1230  
DB 1133 LOOSEBQLRREERKRRQRELOERYAREEBELQROKQRYRDEQSSDLKQWEEFYNA 1192  
DB 1231 ANLRKRYEECEVLYNAKKELSAKD--ENKELHPSIBOETK--QOER-----CONSLT 1261  
DB 1193 VMDKKNVYKGRWEDPQRLQDSQVNDQSDQDLQHLQGLQSDRLQDEERQKQNNHPT 1252  
DB 1282 TITTELQRIQSSAKQIKENKQNTTEILLQDQGLQSDRLQDEERQKQNNHPT 1341  
DB 1253 EEDQLREERQKELKRRDKSGEQFOLLRRREERKRRRQDTDR-----KPREEDLOE 1305  
DB 1342 LQOQVRSLOOLADARQHOEVIAIVTLLSLASGCH-----NDEVOV 1384  
DB 1342 LQOQVRSLOOLADARQHOEVIAIVTLLSLASGCH-----NDEVOV 1384

Thu Apr 11 07:43:31 2002

us-09-750-590-2.std.ra1

Page 17

DB 1306 REOPILROER---DRKPRE-----EELHOGCRKFLBEPORLREERENKTLKEEQ 1355  
OY 1385 ALLQIOMKQ 1395  
DB 1356 LRLEREDLNO 1366

Search completed: April 10, 2002, 18:10:37  
Job time: 110 sec

•  
•  
•  
•

**THIS PAGE BLANK (USPTO)**



alignment\_block:  
08-09-750-590-2.k AP322916  
Align seg 1/1 to: AP322916 from: 1 to: 4428

9 ALeProlysaAmnAsglInAlaAlaAspTrpAniLysTyTrpAspAspArgLe 25  
113 GCGCCGCCGCGCGCATCCGACATTCGAAATTAATATGATGACCCATTT 162  
25 UMeCAspAlaAlaGluAlaGlyLysAspAlaGluLysValSerSerLeuAla 42  
163 GATGAAAGCAGCGGAGGAGGATGTGAGAAAAGGAGCCATCTCTTG 212  
42 LeuLysGlyLysAlaAspProGlyLysLeuAspAlaGluLysArgSerAla 58  
213 CTAAAGAGGGGTAAATCCAGGCAACTAGATGATCGAAGGAGCAATCTGC 262  
59 PheAlaValAlaAlaSerLysGlyLysAlaLeuGluLysLeuAlaLeu 75  
263 TTCCATGCTGTGACCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 312  
75 uLleHleGlyAlaAspAlaLeuThrSerAspPheAlaGluLysArgAlaL 92  
313 TATACATGAGCTGATATTACAAAGCAAGTGCATCGAGGAGGAGGAGGAGG 362  
92 euHleAlaAlaAlaLysTyroLysAlaAlaLeuLysLeuLysLeu 108  
363 TTCACCTGCTGCTACAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 412  
109 GluTyTrpAniAspProGluAlaValAlaAspLeuGluLysArgPheAla 125  
413 CAGTACAAATGTGCTCCAGTACGACAGACCTCGAGGAGGAGGAGGAGGAGG 462  
125 uHleAspAlaAlaMetAlaAspCysProSerSerLysGluLysLeuCysAl 142  
463 TCGAAAAAGAGCAATGCGACATGTGCTCTTCACATACACGCTCTTGTG 512  
142 mphHleGlyAlaSerValAspAlaLysAspAlaGluLysArgPheLeu 158  
513 ACCATGGGCGCTGTGAAATCCACAAAGTGTAAAGAGGAGGAGGAGGAGGAG 562  
159 ValLeuAlaLysGluLysCysAspArgProGluLysGluLysLeuLysAla 175  
563 GTTCTGTGCTACTCAAGTACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 612  
175 PArgGlyAlaAspAlaAspSerArgAspLysGluAlaAspAlaGluLys 192  
613 TACAGAGCCGAGTGTAAATCCAGAGCAAAAGAGGAGGAGGAGGAGGAGGAG 662  
192 euLysGlyCysGluTyroLysLysAspAlaValGluValLeuLysVal 208  
663 TGGTACGTGCTCAATATGCTGACAGATCCAGTACAGTACGATCTTAATTA 712  
209 ArgGluAlaAspValThrLeuLysAlaLeuGluLysAspSerSerTy 225  
713 AATGCTGCTCATATAGCTTGTGATCCCTGGCCATGATGTTCTTA 762  
225 TTYrAlaAlaArgLysLysAspAlaLeuAspAlaLeuThrLeuLysVal 242  
763 CTATGCAACAATGTTGTCACATCTGACATCTTAACCTGTTGATCAAC 812  
242 LysGluLysAlaSerLysArgLysGluLysLeuLysPheLysGlyLysProSer 258  
813 CATCGGAAATATCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862  
259 LeuGluGluAlaTrpAniLeuSerGluLysLeuAlaAspGluLysAlaThrLysSer 275  
863 TTTCGAACGCGAAATTTGACACAGCATGATGCAACATGAAATGTCGAACTC 912  
275 TrpAniGluArgGluLysGluLysAlaLeuLysAlaLeuLysGluLys 292  
913 ACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 962

292 sPheLysGlyLysArgLeuAlaArgLysLysGluGluGluGluAlaGluLysLeu 308  
963 ATTTGAAAGAGAGCTTGAGGAAATATTCAGCAAGAGCAAGGAAATATCTTTTG 1012  
309 AspLysValAspGluLysGluLysGluLysAlaGluLysValMetValAla 325  
1013 GATTAAGCTCAATGTTTACAGTACAGCTGCAAGTACAGGAGGAGGAGGAGGAG 1062  
325 AspAspAlaLeuLysGluLysGlyLysLysLysSerLeuLysAlaAla 342  
1063 TCATCATCTGGAAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1112  
342 yGluLysGluLysGluLysGluLysLeuAlaGluLysAlaLeuLysSer 358  
1113 AAGAAAGCAACATGAAAGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1162  
359 ArgPheLysTyroPheGlyLysArgLysAlaLeuLysGluLysArgPhe 374  
1163 ACATTTAAATTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1212  
375 ... ArgLysGluLysPheLysLeuLysGluLysGluLysGluLysMetTrpAla 390  
1213 TACCCGAAAGAGATGCTCTTAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1262  
390 sPSeGluLysGluLysSerThrGlyLysProValHisMetGluSerArgSer 406  
1263 ACTCAAGTACTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1312  
407 MetLeuArgProLysGluLysAlaLeuProAspGluAlaLysSerTySerGlu 423  
1313 ATGTTAAAGCTCGAATCTTAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1362  
423 uAspGluLysLysGlyLysGluLysGluLysAlaMetArgPheLysPheCys 440  
1363 AAGTAAATTTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1412  
440 ePAlaLysGluLysAspArgLeuLysGluLysAlaGluLysAlaLysVal 456  
1413 CAGCAAAAGCAAGCAAGTAAAGTCCAAATGAGGAGGAGGAGGAGGAGGAGGAG 1462  
457 AluGluLysValAlaLeuAlaLeuGluLysCysGluLysValLysGluLysPhe 473  
1463 GCGAAATGAGAGCTTTAGCATTAAGATGTAAGAGGAGGAGGAGGAGGAGGAG 1512  
473 TrpAspGluLysLysGluLysAspAlaLysLysAspValGluLysAla 490  
1513 AGATGAACAGATTAAGCAATTAAGAAAGTCTTAAAGAGGAGGAGGAGGAGGAG 1562  
490 rGMetTyroLysSerGluLysLysLysLysGluLysGluLysAlaThrHisPheLeu 506  
1563 GAGTATATGATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1612  
507 AlaLeuLysGluLysLeuThrSerAspAlaAlaThrGlyLysAlaSerTy 523  
1613 GCCCTTAAAGAACTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1662  
523 uMeGluLysLeuLysAspAlaLeuLysLysPheLysValLysTyroLysGlu 540  
1663 AACCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1712  
540 LysAlaSerAlaGluLysAlaGluLysAlaGluLysAlaGluLysGluLys 556  
1713 GTGCTTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1762  
557 MetLeuValGluLysPheLysArgAspGluLysLysLeuMetGluLysGlu 573  
1763 ATGATAGTGAAGAGGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1812  
573 uLysArgGluLysGluLysGluLysLeuLysMetGluLysAlaGluLysVal 590  
1813 TAAAGCATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1862

[illegible]





```

|||||
363 TTCACTGCTGCTCTAGTATGACATCATGTGCTCCCTCAAAAACCTTTCA 412
109 GlnTyrAmcProThrGluHisValAspLeuGlnGlyArgThraIala 125
413 CAGTCAAAATTTGCTCCAGTACACATCCAGCTACAGGAGAAAGACGACT 462
125 uHisAspAlaIaIaMetAlaAspCysProSerSerIleGluLeuGlyAsp 142
463 TCAAAAAAGCAATGGCAGATATGCTCTTACGATACATACGCTCTTGTC 512
142 spHISgIyAlaSerValAsnAlaIyAspValAspD1YAgrThProLeu 158
513 ACCATGGGAGCTCTGTGATATGCAAGATATGAAAGAGGAGCAACACT 562
159 ValLeuAlaThrGlnGlnMetCysArgProThrThrIleuAlaHis 175
563 GTTCGGCTACCTCAGATAGTACGCCAACAAATATGCAACTGCTGTATA 612
175 pArgIyAlaAspIleAsnSerArgAspIyGlnAsnArgThraIalaLeu 192
613 TAAAGAGGGAGATGTAATATTCAGAGCAAAACAAACGAACTGCCCTTA 662
192 eIleuGlnGlyGlyTyrGlyGlyAspAlaValGluValLeuIleLys 208
663 TCGTAGCTGGCAATATGCTTCCAGAAATGCAATAGAGCTTAAATTA 712
209 AsnGlyAlaAspValThraLeuAspAlaLeuGlnYHisAspSerSerTy 225
713 AATGGCTGATATTAAGCTCTCGATGCTGCTGGCCATATAGTCTTAA 762
225 rTyAlaArgIleGlyAspAlaLeuAspIleLeuThraIleuLysThra 242
763 CTATCGAAGAAATTCGACAACTCGACATCTACGATCTGTGAAAGCTG 812
242 IAserGlnAsnSerAsnIyGlyArgGluLeuThrLysGlyAspProSer 266
813 CATCGCAAAATTCGCAAAAGCAAGCAACTTGGAGAAAGAAAGCACTAT 862
269 LeuGlnGlnArgAsnLeuSerGlnMetLeuAspGluValAsnThrLysSe 275
863 TTGCAGACGCAAAATTTGACACATCTGACATCTGAAAGTGAAGTGC 912
275 rAsnGlnArgGluHisGlnAsnIleGlnAspLeuGlnIleGlnLysGln 912
913 ACATCAGAGGAGCATCAAAATATTCAGATTTGGCAGATTGCAAAATGAG 962
292 spLeuLysGlnArgLeuArgLysIleGlnGlnGlnIleArgIleLeuLeu 308
963 ATTTCGAAGAGATGTCGCAAAATATTCAGCAAGCAACGAAATGCTTAA 1012
309 AspLysValAsnGlyLeuGlnIleuGlnIleuAsnGlnGlnIleuValMet 135
1013 GATTAAGTCATATGCTTACAGTTACAGCTGAAGCAAGCAATTAATGCT 1062
325 aAspAspLeuGlnIleuSerGlnLysGlnLysLeuLysSerLeuLysAla 342
1063 TGTATGATCTGGAACGAGAGCAAGAAAGGCTGATCCCTTTGGACGCTA 1112
342 YSGGlnLysGlnHisGlnIleuSerLeuArgThraIleGlnAlaLeuLys 358
1113 AAGAAAGCAATCATAGAAACCTTAAAGCATATTCAGAGCTCTGAAAT 1162
359 ArgPhyLysTyrPheGlnIleuSerAspHisLeuGlnIleSerGlyHis 374
1163 AGATTTAATATTTGAGAGTGCATCATTAAGTACGAGAACTTCATTCG 1212
375 ... ArgLysGlnAspMetLeuLysArgGlnGlnIleuGlnMetCysPhe 390
1213 TAAACCAAAAGAAAGATATGCTCTTAAACAAGGTCAGATGTATATGCGAG 1262
390 spSerGlnCysThrIleSerThrglyMetProValIleMetGlnIleArg 406
|||||
|||||
1263 ACTCAAGCTGACTTCCCAAGCTATACCAAGCCCATATTCAGCAAGCAATCT 1312
407 MetLeuArgProLeuGlnLeuAlaLeuAlaLeuProAsnGlnIleAspSer 423
1313 ATGTTAAGACCTCTGAAATCATCTTTACCCAGCTCAAAAGCTATACCTGTA 1362
423 uAsnGlnIleLeuLysGlnIleuGlnIleuAlaIleuLysArgPheCysAsp 440
1363 AATGAAATATTTTAAAGAAAGAGCTTTAAGCAAGCAATGCAATCTTGAGAT 1412
440 eRAIleLysGlnAsnArgProLeuLysLeuGlnIleuAlaIleuVal 1466
1413 CAGCAAAACCAAGACCACTGAACTCCAAATTAATATCTGGACACAAAGTG 1462
457 AlaGlnCylsValAlaLeuAlaLeuGlnLysGlnIleuValLysGlnAsp 473
1463 GCGAATATGCAAGCTTTAGATTTAGATTTGAATGAAGAGCTGCAAGAGATTC 1512
473 rAspGlnGlnIleGlyGlnIleuGlnIleuValLysAspValGlnLys 1562
1513 ACATGAAACGATTAAGCAATTAAGAAATGCAATTAAGATATGTCGAAAG 1562
490 rGlnMetLysGlnIleuSerGlnLysValLysGlnMetGlnIleHisAsp 506
1563 GGATGATGATGATCAGAGCAATTAAGCAATTAAGCAATGCAAGCAATTTCTT 1612
507 AlaLeuLysGlnIleuLysThrSerAspAlaIleuIleGlnYAsnHisArg 523
1613 GCCCTTAAAGAACTTAAACAAATGAGAGAGCTCTCAGAGAACTACAGACT 1662
523 uMetGlnGlnLeuLysAspGlnLeuLysAspMetLysValLysArgGln 540
1663 AACCAGGAAGCTGAGAGATCAGTTGAAAGCACTTGAAAGTAAATATGAG 1712
540 YAlaSerAlaGlnValGlyLysLeuArgAsnIleIleLysGlnAsnGln 556
1713 GTGCTTCAGACAGATGGGAAATTAAGAAACCAAAATGCAAAACAAATGAG 1762
557 MetLeuValGlnGlnIleuLysArgAspGlnLysGlnMetGlnLys 573
1763 ATGATGATGAAAGACTTAAAGAGGATGAAGCAAGCATGATAGCAAGAAA 1812
573 uLysArgLeuGlnIleuLysGlnIleuSerMetCysGlnIleuGlnIleu 590
1813 TAAAGCTATTCAGAAAGATCTGATGTGTAAGTAAAGAGCAAGCAAGCA 1862
590 rGAGIArgGlnLysLeuThrglyMetGlnGlnIleuLysArgLeuSer 606
1863 AAGGAGAAAGGTCACAGATGGAAGCCAGCAAGCAAAATATGTACAG 1912
607 LysLeuAlaLeuSerIleProAlaGlnLysArgPheGlnLysMetLysSer 623
1913 AAGTTGGCCCTTCATTCACGCTGCAAAATTTGAAAGCAAGCAAGCTC 1962
623 uLeuSerAsnGlnLeuAsnGlnLysValLysGlnIleuIleAspValGln 640
1963 ATTTCAATGAAAGCAATGAGAAACCAAAATATTCGTAAGAAATGAGAA 2012
640 rGlnLysArgLysSerLeuAsnGlnIleuThraIleProLeuLysArgGln 656
2013 GAGAGATGAAAGAAATCATCTTGAATTAAGCAATTAAGCAAGCAAGCT 2062
657 GlnAsnArgLysValLysLeuAlaGlnHisGlnIleuLysProGln 673
2063 GAGATGTTAAGCCAAAGCTTGGTCAGCAAGCTCAACAGAGAGCAATG 2112
673 uGlnLeuLysSerArgLeuGlnGlnLysSerGlnGlnIleuGlnLysArg 690
2113 ACAGGTTAAGCAAGATTTAGAAACCAAAATCAGAGAAATCTCCGAAAGAG 2162
690 IeThrGlnLeuLysSerLysArgGlnIleuGlnLysGlnIleuGlnLys 706
2163 TCACATGATTAACATTAAGAAATTCAGAACTCAAAAGCAATTAATTAAG 2212

```

[illegible][illegible]

```

|||||
4013 ACTGAGCTCTTAATGATGTCGAAAGATTAAACAGCCACTGACAGGCT 4062
1321 userGlnmetHrTgLyserGlyserProserLyserGlnInserGln 1338
|||||
4063 TTCGCACTGACCTTACCAAGTGGAGCCCAAGCCAGACAGCCAGC 4112
1338 enUlaSerLeuGlnInGlnValaHsSerLeuGlnInGlnLeuLa 1354
|||||
4113 TGATTGACACTGTCAGCCAGCCAGGAGAACTCTGCGAGAGCGCTGCG 4162
1355 AspaLaAspArgInHsGlnGlnValaIleAlaIleGlyArgHsHisLe 1371
4163 GATGCTACAGCAGCAGCAAGCAAGTAATTCGCAATTCGCGACACACT 4212
1371 uGnSerAlaInGlnGlnHsMetAspGlnAspValaGlnAlaIleLeu 1388
4213 TCTTAGTGTGCGACAGGTCACATGATGATGAGTTCGAGAGCTCTGCG 4262
1388 euGlnIleIleGlnMetArgInGlnIleuValaCys 1399
|||||
4263 TCGCAGATCATCAATATGCGCAGCGGCTTGTCGTCG 4298
seq_name: gp_om:CFCS3VS

seq_documentation_block: 4643 bp mRNA MM 20-OCT-2000
LOCUS CFCS3VS
DEFINITION Canis familiaris mRNA for C3VS protein.
ACCESSION X99145
VERSION X99145.2 GI:10944717
KEYWORDS C3VS protein.
SOURCE dog.
ORGANISM Canis familiaris
Bukharaia Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Carnivora: Fissipedia: Canidae: Canis:
1 (bases 1 to 4643)
Mullin,F., Savonet,V., Radulescu,A., Petermans,J., Dumont,J.E. and
Moenhau,C.
Identification and characterization of novel genes modulated in the
thyroid of dogs treated with methimazole and propylthiouracil
97067069 Chem. 271 (45), 28451-28457 (1996)
2 (bases 1 to 4643)
Savonet,V.
Direct Submission
Submitted (06-JUL-1996) V. Savonet, Institute of Interdisciplinary
Research - UB, Building C (local C4-124), Campus Erasme, Route de
Lemik 808, 1070 Brussels, BELGIUM
3 (bases 1 to 4643)
Savonet,V.
Direct Submission
Submitted (20-OCT-2000) V. Savonet, Institute of Interdisciplinary
Research - UB, Building C (local C4-124), Campus Erasme, Route de
Lemik 808, 1070 Brussels, BELGIUM
On Oct 21, 2000 this sequence version replaced g1:1429313.
Location/Qualifiers
organism="Canis familiaris"
/db_xref="taxon:9615"
/tissue_type="thyroid"
/clone_id="lambda ZAP11"
328..4494
/gene="C3VS"
328..4494
/gene="C3VS"
328..4494
/notes="overexpressed in thyroid tissue after TSH
stimulation"
/codon_start=1
/product="C3VS protein"
/protein_id="CAC14169.1"
/db_xref="GI:10944718"
/translation="MPSLLLATRNQIISMAKNCWFCAPKRRHADNKNKYDRLMKAA

```

```

BRGDVEKVSIIAKKGINFGKLDVEGRSAFHVAVSKNGLELAAIILINGVITTTSDTA
SNKILIAKRYGALICLQKLILOYKCPTEHADLOGRTALMDAAADCPSSIQLLDNDHA
GNAKIVKQRPRLVAVDNGRRAICOLLIRNAGEINSRKQNTKMLAGCEGCKDAV
GNDVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLV
RILPYVAVDLSLVLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLV
BEVAVVADLSSEKELKSLIAKPKRQHESELPTEISL KQDQSLILKRGSLQDLS
SMLEPLETEISLPNQTSENDLLKKELEAMRTCESEAKQORIKLQNTLALHVAKDS
LECRALKEKSDQDIKQEDALIKVDOKRYSSGCKYKQKQNTFRALKKEILTSKALISG
RMEKEIKQIKQKMKAKYEGASAEVGLKANOIKQNTLESLYPEFSKSLSSLEVEKVKIG
LSMCEITBOKKGRVAAVEGQVKBALATLTLSPLEKFSKSLSSLEVEKVKIGE
TERSEKSLTEIRDLRELEKCAKLAQVAPREBOLSKLDKSKSMDIVTEBELKOL
KNQVLDQVREKATVLDKLMQDVHNTSEIKSHVPLQVSEKSKSHDVTBELKOL
KNQVLDQVREKATVLDKLMQDVHNTSEIKSHVPLQVSEKSKSHDVTBELKOL
DILKPKQGDDEKNTNLSQVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLVSLV
DARKKKEBLODGEFTVKTQDNEILKRLKLENTQSLKAEVTLSPHEKESPA
QDNSEALILAVYRKQGEIYTLDAIEADKKEIDTIOECIKLYAVPISFEDEKSPA
TEKLEKQLESDQKRYHVESEAKKYKQEDMKKELEPTLOKDKKXVLEIKSHDIE
BALNRAREILNKQLKDLQKYSERKTERKELVDONARQTSLLAQDTLOKQHPYLE
VEULKRSJSTTEILKELKPKQCYEKEDQTVAKLHQLKQKNSVPLGSHLRYVE
ALFEKQNTSLASKEEESQNTTEVSKLOSERVDQTKQALQKLETRPVEDLSYKAT
BROKESLQVLSLQSLKELKELKELKELKELKELKELKELKELKELKELKELKELK
KRSKSLQIDTLOHVKSLDQKLDQDNDQRIEYVILNITLILSLAAQDMDVQALLO
ITQKRGVC"
BASE COUNT 1645 a 872 c 1185 g 941 t
ORIGIN
alignment_recres:
  QmRatio: 4.523 Length: 1401
  Percent Similarity: 95.575 Percent Identity: 84.654
alignment_block:
  US-09-750-590-2 x CFCS3VS
Align seg 1/1 to: CFCS3VS from: 1 to: 4643
1 MethSerGyTgPheSerCyGAlaPArgLyAspArgGlnAlaIleLaas 17
|||||
373 ATGATGACACTGTTGGTCTTTCTTGCTGCTCCCAACAACAGCATGACAGCA 422
17 pTPAsnLysTyrAspAspArgLyMetLysAlaIleAgLysGlyAspV 34
|||||
423 TTGGCAACAAATATGATGATCCCAAGTGTGAAGCCCGGAGGAGGAGTGG 472
34 aTGlucyGValSerSerTleuAlaIleuLysGlyGlyAspArgGly 50
|||||
473 TGGAAAAGTTTCTCAATCTTCTGTTAAAGGGCATCATCGACGCAAA 522
51 LeuAspValGlnLysGlyAlaPheHsValaIleSerLysGlyAs 67
|||||
523 CTACATATGGAAGGACATCTCCCTTCATGATGTTGGCTCCAAAGGGGAA 572
67 pLeuGlyCysLeuAspAlaIleuLysHsGlyValaAspLleHrThrs 84
|||||
573 TCTTCATATGTTTAAATGTCATCTTATCATGATGATGANTATCAACA 622
84 erAspThrAlaGlyTyrAspAlaIleuLysHsValaIleValaIleGlyAsp 100
|||||
623 GTGACATGCGAGGAAGAAATGCTCTTCATCTGTCGTCGCAAGATGCGCAT 672
623 GTGACATGCGAGGAAGAAATGCTCTTCATCTGTCGTCGCAAGATGCGCAT 672
673 GCATTGTGTCACAAAACTTCTACAGTAAATATGTCCTCCCATGACAAAC 722
101 AlalaGlyCysLeuGlnIleuLysGlnGlnIleuLysGlnProHrThrs 117
|||||
117 LAspArgInGlyArgThrAlaIleuLysAspAlaIleMetLalaGlyAsp 134
|||||
723 AGACCTGACGAGGAAGAAACCGCACTTCATGACGCGACCAAGGCAAGATGTC 772
134 rGserSerTleuIleuLysGlnGlnHsGlyAlaSerValaIleAlaLys 150
|||||
773 CTTCGACATGACGCTCTTTCGACATGAGGCGCTCCGTATGATGCAAA 822

```

151 AspylaAspylaATGATGThcProleuValLeuAlaThcGlnMetCysArgPr 167  
 167 cthrtIcYsGlnLeuLeuIleAspArgValAlaAspIleAsnSerArgA 184  
 873 AGATATCTGCACACTCTGATATGATCAGGGGCGAGGATTAATCCAGAG 922  
 184 bplYsGlnAsnArpThAlaLeuMetLeuGlyCysGluTyrGln 200  
 923 ACMAACMAACGAGACTCTCATCTCTGGTCCAGATGGCTGTAG 972  
 201 AspAlaValGluValLeuIleLysAsnGlyAlaAlaAspValThrtIleLeuAs 217  
 973 GATGCTGTAGAGCTTACTTAATAAATGGCTGATAGTACCTCTGCGA 1022  
 217 palLeuGlnYHlaAspSerSerTyrTyrAlaArgIleGlyAspAsnLeuA 234  
 1023 TGCCTTGGGCGATGATAGCTTACTTACTTACTTACTTACTTACTTACTT 1072  
 234 bplIleuThrtIleuLeuLysThrtAlaSerGlnAsnSerAsnLysGluYArg 250  
 1073 ACATCTTACTTATTTGATGACATGGCTGCAAAAATACCAACMAAGAGAG 1122  
 251 GluLeuTrpLysGlyProSerLeuGlnGlnArgAsnIleuSerGlnMe 267  
 1123 GACCTTGGAGAGAGAGACACTCTTACACAGCAAAATTTGGCGCTAC 1172  
 267 LLeuAspLysValAsnThrtLysSerArgAsnArgGluLysIleAsn 1162 284  
 1173 GCTAGATGATGAATGATGATGATGATGATGATGATGATGATGATGATG 1222  
 284 LAspLeuGluIleGluAsnGluAspLeuLysGluArgLeuArgLysIle 300  
 1223 AGGAGCTGGAGATTAATGATGATGATGATGATGATGATGATGATGAT 1272  
 300 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 317  
 1272 CACGACAGACAGATATTACTGATTAAGCTCAATGGTTTACACTACA 1322  
 317 nLeuAsnGluGluValIleValAlaAspAspLeuGlnSerGluYArgLys 334  
 1323 GCTGATATGACGAACTGCTGCTGATGATGATGATGATGATGATGATG 1372  
 334 yLeuLeuLysSerLeuLeuAlaLysGlnGluYArgLysGlnSerLeu 350  
 1373 AGCTGAATCTCTTGTGGCTTAAGAAACACACTGAAGAAAGCTTA 1422  
 351 ArpThrtIleGluLysSerArgPheLysTyrPheGlnLysSerArgP 367  
 1423 AGAAGCTATGAGCTCTGAAACACAGATTTAAATTTTGGAG 1464  
 367 nLeuGlySerGlySerHisPheArgLysGluAspMetLeuLeuLysGln 384  
 1464 ..... 1464  
 384 LysGlnMetYrMetThrsAspSerGlnCysThrsSerThrtYrMetProVal 400  
 1465 ..... TGTACTCTCCAGGGGCTGGCCAGCC 1488  
 401 HisMetGlnSerArgSerMetLeuArgProLeuGluLeuAlaLeuProAs 417  
 1489 CACATCCGACAGAGCTCATGTATTAACCCCTGAGCTGATCATCCACCA 1538  
 417 nGlnAlaSerTyrSerGluAsnGluIleLeuLysGlnGluLysGlnLys 434  
 1539 TCMAACCTCATATCTGAAAATGACCTTTAAAGAAAGAGTTAGAACAA 1588  
 434 eCArgThrPheCysAspSerAlaLysGlnAspArgLeuLysLeuGlnAsn 450  
 1589 TGAGAACTTTCTGCGATTCAGCAACAGAACCCGGCTCAGCTCCAGAAC 1638

451 GluLeuAlaHisLysValAlaGluCysLysAlaLeuAlaLeuLysGlu 467  
 1639 GAGCTGGCCACAGAGCTGGCTAGCTGCAAACTTATAGACTGTAGAGTGA 1688  
 467 nArgValLysGluAspSerAspGluGlnIleLysGlnIleGlnIleAspAla 484  
 1689 ACCATCTAGAGAGACCTGTATGATGACAGAGCTGATGATGATGATGAT 1738  
 484 eLysAspValGlnLysArgMetGlyLysSerGlnGlyLysValLysGln 500  
 1739 TCMAAGATCTCCAGAGAGAGAGATGATGATGATGATGATGATGATGAT 1788  
 501 MetGlnThrsPheLeuAlaLeuLysGluLysLeuThrsSerAspAlaVal 517  
 1789 ATGACAGACACTTCTTGGCTTAAAGAGCCCTGACAGAGTGAAGACAG 1838  
 517 ArpThrtLysAsnHisArgLeuMetGlnGluLysLysArgPheLysArgP 534  
 1839 TATAGGATTCAGACACTTATGATGATGATGATGATGATGATGATGAT 1888  
 534 eLysValLysTyrGluGluValIleSerAlaGluValGlyLysLeuAsn 550  
 1889 TGMAAGGAAATTTGAGGGCTGATCAGACAGAGCTGAGAAATCCGCAAC 1938  
 551 GlnIleLysGlnAsnGlnMetLeuValGluLysPheLysArgPheGlu 567  
 1939 GAAATCTGACAGATGATGATGATGATGATGATGATGATGATGATGAT 1988  
 567 YLysLeuMetGlnGluAsnLysArgLeuGluIleLysGlnLysSerMetCys 584  
 1989 CAGCTGCTGACAGAGATTAAGCATTTGACAGAGAACTCATGTATGCG 2038  
 584 LLeuGlnArgGluLysArgGlyArgLysLeuThrtGlnMetGlnGln 600  
 2039 AAACGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088  
 601 LeuLysAspLeuSerAlaLysLysLeuSerIleProAlaGluLysArg 617  
 2089 GTMAAGGACTTCTTAAAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 2138  
 617 eGluAsnMetLysSerLeuLeuSerAsnGlnLysLeuAlaLysLys 634  
 2139 TGAGAGCATTAAGAGCTTATTAACAGACAGATTAATTAATTAATTAAT 2188  
 634 yLeuLeuIleAlaValGlnArgGluLysArgSerLeuAsnGlnThrsArg 650  
 2189 AATTTGATGAGACAGAGAGAGAGATTAATTAATTAATTAATTAATTAAT 2238  
 651 ProLeuLysArgGluLysGlnLysAsnLysAlaLysLysLysLysVal 667  
 2239 CAGTTAAAGAGAGAGCTGAAATTTGAAAGCAAACTTCTCTGACAG 2288  
 667 LysPheGluGluIleGluLysGlnLysSerArgLeuGlnGlnLysSerG 684  
 2289 CMAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2338  
 684 LysIleLeuGlyLysArgIleThrtGluLeuThrtSerLysAsnGlnThrsLeu 700  
 2339 GAGAACTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2388  
 701 GlnLysGluIleGluLysValCysLeuAspArgLeuGlnGlnGlnGln 717  
 2389 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2438  
 717 nValAsnAsnLeuThrtThrtGluMetLysAsn.....ValProLeuLysVal 732  
 2439 AGTACATTAATTAACAGAGTAAATTAAGAGATTAATTAATTAATTAAT 2488  
 732 nSerGlnGlnMetLysSerHisAspValIleValAspAspLeuAsn 748  
 2489 TGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2538  
 749 LysLysLeuSerAspValThrtHisLysTyrThrtLysLysLeuGlnLeu 765





```

490  rMetyrGluSerGluGlyValValGluGlnMetGluThrHisPheLeu 506
    |||
2661  GATGTATGATGACGAGAGCTTAAGCTTAACAAATGACACACCATTTCTT 2710
    |||
507  AAlaLeuGluGluHisLeuThrSerAspAlaIaIaThrGlyAsnHisIaArgLe 523
    |||
2711  GCCCTTAAGAAACACCTTAACAGAGTGAACAGCGCTCCAGGAAATACAGACAT 2760
    |||
523  uMetGluGluLeuGlyAspGluLeuGlyAspMetGlyValIleGlyTyrGluG 540
    |||
2761  AACCGAGAAAGCTGAGAGGATCAGCTTAAGACCTTAAGAAATTAATTCAG 2810
    |||
540  lYAlaSerAlaGluValGlyLysLeuArgAsnGlnIleGlySerGlnAspGlu 556
    |||
2811  GTCCTTGACCGAAGAGGGGAAATTAGAACAAACCAATCAACAAATTAAG 2860
    |||
557  MetLeuValGluGluHisPheLysArgAspGluGlyGluMetGluGluGlu 573
    |||
2861  ATGATAGCTGAAGACGTTTAAGAGGATGAAGGACGAACTGATAGAGAAA 2910
    |||
573  nLysArgLeuGlnIleGlyLysLeuSerMetCysGluLeuGluIaArgGluVal 590
    |||
2911  TAAGCGATTACAGAAAGAGACTTACATGATGTGTAATTAAGGACGAGGAA 2960
    |||
590  rGGLyArgGlyLeuThrGluMetGluGlyGluLeuLeuLysLeuLysSerIaIa 606
    |||
2961  AACGAAAGAAAGCTCAAGACATGAGACGACGACGACGAAAGAAATGTGACG 3010
    |||
607  LysLeuAlaLeuSerIleProAlaGluLysPheGluAsnMetLysSerIle 623
    |||
3011  AAGTTGGCCCTTCCATTCACGCTGAATAATTTGAAACATGAGACGCTC 3060
    |||
623  uLeuSerAsnGluLeuAsnGluLysAlaLysValLysLeuIleAspValGlu 640
    |||
3061  ATTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3110
    |||
640  rGGLyTyrGluArgSerLeuAsnGluLysArgPheGluValArgGluLeu 656
    |||
3111  GAGAACATCAAAAATCATTAGCTGAATTAAGACAGTTAAAGAGAACATT 3160
    |||
657  GluAsnLeuLysAlaLysLeuAlaGlnHisValLysProGluGluHisGlu 673
    |||
3161  GAGAACTGTTAAGCGACAGCTGCTGACACGCTCAACGACAGAGAACTGA 3210
    |||
673  uGluLeuLysSerArgLeuGlnGluLysSerGlyLysArgGluLysArg 690
    |||
3211  ACAGCTTAAGAGCGATTAGACAGAAATCAAGAGAACTTGAGAGAA 3260
    |||
690  lThrGluLeuThrSerTyrAsnGlnIleThrLeuGluGluGluIleGluLys 706
    |||
3261  TCACGAGATTAAACATTGAAAAATTCAGACATCAACAAAGAAATGAAAA 3310
    |||
707  ValLysLeuAspAsnLysLeuLeuThrGlnGluIleValAsnAspLeuThr 723
    |||
3311  GTTTAATTGATTAAGCTCTCAAGAGACGACAACTAACTAACTAACTAACT 3360
    |||
723  rGluMetLysAsn.....ValProLeuLysValSerGluGluMetLysL 738
    |||
3361  TGAATAAGAAAAATCATATGCTCTTAAAGTAAGTAAGACATGAAAA 3410
    |||
738  ySerHisIaSerValIleValAspAspLeuAsnLysLysSerAspVal 754
    |||
3411  AGTCACATGATGCAATGATGATGATGATGATGATGATGATGATGATGATG 3460
    |||
754  ThrHisLysTyrThrGluLysLeuGluMetGluGluLysLeuMetGlu 771
    |||
3461  ACACAAAAATATACAAAAAGAGCTGGAATATGGAATAATCTCTACGGA 3510
    |||
771  uAsnAlaSerLeuSerTyrAsnValSerArgLeuGluIleThrValPheIle 788
    |||
3511  AAGTACAGCTTAAAGTAAGATGTAAGCGCGCTGAAGAACTGTGCTTCAC 3560
    |||
788  rProGluArgHisGluLysGluMetAlaLeuLysSerAsnIleThr 804
    |||
    |||
3561  CTCGAGAAACATGAAAGAAAGAAATAGCTGCAAGAAATTAATGTT 3610
    |||
805  GluLeuLysAlaGluLeuSerGluLeuAsnLysLysCysGluGluValAsp 821
    |||
3611  GAACCTTAAGAAACAGCTGTCTGAACCTTAAGAAAAATATGCTGTGAAGACA 3660
    |||
821  nGluLysIleTyrSerLeuMetSerGluAspAsnAspLeuLysTyrThr 838
    |||
3661  GAGAAAGAAACAGCTCTCACTGCAAGAAACCAATCTTCAAGAAAGATCA 3710
    |||
838  eTSerHisGluIleThrValProValIleThrHisGluIleLysThrAla 854
    |||
3711  TGAATATCAGATGATGCGCAATTAACCCATGAAAGATTAATAATGACA 3760
    |||
855  LeuSerSerThrLeuAspLysTyrHisAspArgGluLeuValIleAspValLys 871
    |||
3761  CTGAATGATGATTAAGCCCAACCAATCAAGACAGATTTTAAAGATGluSer 888
    |||
871  sLysCysGluAspIleAsnGluLysPheValIleLysAspArgLysSer 896
    |||
3811  AAATTTGAAGATTTAAATCAAGAAATTTTAAATTAAGATTAAGATTC 3860
    |||
888  lLeuLeuArgAsnLysGluValAsnThrGlnAsnIleValLysAlaGlu 904
    |||
3861  AAATTAATTAAGAAACCTGAAAAACCTCAAGAAACCAATTAAGAAACCTGAG 3910
    |||
905  ThrIleSerLeuArgGluHisGluGluLysMetSerGlyLeuAsnArgLys 921
    |||
3911  TACATCAGCTGCGAAGACACAGCAAGAAATTAAGCTGCTTAAGTCAG 3960
    |||
921  rMetLysLysValGlnAspAsnSerAlaGluIleLeuAlaLysTyrLysL 938
    |||
3961  CATGCAAAAGCTGACAGATGATTAATGCTGAATCTTGCCCACTACAGAA 4010
    |||
938  ySerSerGlnGluLysIleValThrLeuHisGluGluIleAlaIaLeuLys 954
    |||
4011  AAGGCAAGAAAGATGTGACACTGACGCAAGAAATTAAGCAAGAAAG 4060
    |||
955  ArgGluLeuAspThrIleGlnGluLysIleLysLeuLysTyrAlaProIle 971
    |||
4061  AAGGAGCTGACACAAATTAAGAAATGCAATTAAGTAAATATGCCCAAT 4110
    |||
971  eIleSerLeuGlnGluCysGluValArgPheLysAlaIaThrGluLysGlu 988
    |||
4111  TGTCAAGCTTGAAGACTGCAAGAAATTAATTAAGCAAGAAAGAAAG 4160
    |||
988  euLysGluLysLeuSerGlnGluIleThrGlnLysTyrAsnThrSerGluGlu 1004
    |||
4161  TAAAGACCAATTAAGACAGCAAGCAAGAAATTAAGTATGCTGATGAA 4210
    |||
1005  GluAlaLysLysCysGlnGluAsnAspArgLysLeuLysGluIleLe 1021
    |||
4211  GAAGTCAGAAAGAAACAAACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 4260
    |||
1021  uThrLeuGlnLysAspLeuAspArgValHisIleGluAsnSerT 1038
    |||
4261  TACCTCTCAAGAAACATTTGAGAGAAATTAAGCAAGCTTCATTAAGAGCTC 4310
    |||
1038  yTyrGluThrGluValArgAlaLeuSerArgTyrThrGluLysLeuAspArg 1054
    |||
4311  ATGAAATGCAAAAGCAATTAAGCAAGAAATTAAGCAAGCAAGCAAGCAAG 4360
    |||
1055  LeuLysAspLeuGlnGluLysTyrThrGluAlaLysLysGluLysGlu 1071
    |||
4361  TTAAGAACTTGTCAAGAAATTAAGCAAGAAATTAAGCAAGAAATTAAGCA 4410
    |||
1071  sLeuValGluLysAlaLysGlnIleThrSerGluLysLeuAlaIaLeuIle 1088
    |||
4411  CCTAGTACGAAGAAATGCGCAAGCAAGCTTCAAGATCTGCAAGCTGCAAA 4460
    |||
1088  hTrpLeuGlnGluLysGlnHisValProLeuGlnGluIleValGluLys 1104
    |||

```





```

alignment_scores:      Length: 952
                        Quality: 3989.50
                        Ratio: 4.433
                        Percent Similarity: 94.338
                        Percent Identity: 83.508
alignment_block:
US-09-750-590-2 x AC087699/rev ..

```

352 GlnCysHisSerThrGlyMetProValHisMetGlnSerArgSerMetLe 408  
|||||  
128284 CAGTGTACTTCCCCAGGTATACCAAGCCCATATGCAAGCAGATCTATATGTT 128295

408 uArgProLeuGluLeuAlaLeuProAsnGlnAlaSerTyrSergLysG 425  
|||||  
128234 AAGACCTGTGGAACTATCTTTACTCCAGTCGAACGCCTCATCCTCAATTG 12901

425 IuIeLeuLysGluLeuAlaMetArgThrPheCysASPSeRa 441  
|||||  
128184 AAATTTTAAAGNAGAGTAGAAGCAATCGCACTTCCTGTGACGTACGA 128135

458 128134 AACACAGACCGACTGAGGCTCCAAATGAACTGGCACACAAAGTGGCAGA 128085

475 lucinilcyscinleugluaspalauleulysaspvalginlysarrmet 491  
128084 ATGCAAGCTTTAGCATTACATGTGAAGGGTCAAGGAGGATTCAGATG 128035

126034 AACGATAAAGCAATTAGAGATGCATTAAAGAGTGTCCAGACAGGATGC 127985

492 TYRGLuserGLGLyValLysGlnmetGlnThrHisPheuAlaLe 508

508 uLysGluHisLeuThrSerAspAlaIatHrGlyAsnHisArgLeuMetG 525  
|||||  
127934 TAAAGAACACTTAAACAAGTGAAGCGACCTCAGGGATATCAAGACCTAACCC 127995

525 IUGLuleuLysASPGLInleuLysASPmetLysValIstYrcIGLyAla 541  
|||||  
|||||

342 SerAAGuVAIGIlylsLeuArgAsnGlnIleLysGlnAsnGluMetLe 558  
|||||  
127834 TCACGAGAGTGGGGAATTAGAAACCAATCAACAAATGAGATGAT 127785  
|||||

558 uValgIugLrPhelYatRAspcJugLYLysLeuneCtUcLIkAsnlysa 575  
:|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||  
127784 AGTGAAGACTTTAAAGGATGAAAGCGCACTCATACAGAAAAATTAGC 127735

592 Arglysluethrclluefllcrl of aut uctocl ouf flltst 500  
|||||.....|||  
127734 GATTACGAGNAGACTTACTATCTGTGGAAATGGACCGACGACAGAAAAAGA 127685

127684 AGAAGCTCACACAGATGGGAAGGCCAGSCAAAAGATTGTCAGCGAACTT 1276535  
608 uAlaLeuSerIleProAlaGluIlyPheGluAsnMetIlySerLeuLeuS 625

625 *era*sngluleauasnglulysalalysylleuileaspvalgluarqglu 641  
|||||:::|||||  
127034 GGCCTTTCACATCCAGCGGAAAAATTGGAAACATGAAGCGCATTAT 127585  
127034

[illegible]

```

658 mteuysalalyseuclaaaglnhisvalysprogluglnhsgluglnl 675
      ::
127484 TGTTAAGGCCAAAGCTTGTCTCAGCAGCTCAACCAAGAGGAAATGACGAGG 127435

```

[illegible]

127384 GAGTTAATCATTTGAAAAATTCAGACACCTTCACAAAGCAATTTGAAAAAGTTTA 127335

725 etlvsash.....ValProleulysValserclnclutlmetlslslsSer 739  
||||||| ||||||| :|||:|||||  
127354 TTTGAAATATAGCTCTCAAGSGAGCAACACTACTTAAACATTGGAA 127285

740 HisAspValIleValAspAspLeuAsnLysLysLeuSerAspValThrH<sub>i</sub> 756  
|||||:::|||||:::|||||:::|||||:::|||||:::  
CATATCAATTATGATGTCTTATATAAAGCTTTTAGAACTAACACA 127185

756 slygTtThrGluLysLeuGIunKcglubLysLeuMetGIuaSna 773  
:::|||||  
227184 AAATATACGAAAGACTGGAAATGTGAAGAATTCTACTGAAAATG 127135

```

790 G|HARAHG|ui|vsc|knot64|TfuuACTCTAATTAAGTTAACTCCT 127085
      |||||:::|||||gucstgumrvalrlllrrtfto /89
27134 ACAGCTTAGTAGCATGTAAACCCGCTGAAAAGTGTTGGTACTCCT 127085
      |||||:::|||||
```

27084 GAGAAACATGAAAAAGACATATTAAGTCTCGAAGATCCAAATTGTTCACCT 127035  
806 uLysLysGlnIeuSerGlnIeuAsnLysLysCysGlyGlnIuapGlnGlu 823

823 yslIeYrSerLeuKetSerGluasnaGnasplLeuLysLysThrMetSer 839  
|||||:||||| |||||:|||||:|||||  
|||:::||||| |||||:|||||:|||||  
|||:::||||| |||||:|||||:|||||  
|||:::||||| |||||:|||||:|||||

[illegible]



13438 AGAAGTCGACAGATGACGAGCGGACGAAAGCAATTCGACGACCT 134389  
 608 |||||SerrLeuPheGluPheGluValMetSerLeuMetLeu 625  
 134388 GGCCTTTCATTCACGCTGCAAAATTTGAAACATGAAAGCGCTATTAAT 134339  
 625 eTASGluLeuAsnGluValLeuValSerValGluValGlu 641  
 134338 CAATGCAATGAAATGAAAGCAAAATATGTAACAAATGCGAAGCA 134289  
 642 TTCTGAGSerrLeuAsnGluPheArgProLeuValArgGluVal 658  
 134288 CATGAAATTCATTAAGTGAATTAAGACAGTTTAAACAGCAACTGAGAA 134239  
 658 nLeuValLeuValLeuValLeuValValSerGluGluLeuGlu 675  
 134238 TGTTAAGCGCAAGCTTCGACGACGACCAACGAGCAACATGACAGC 134189  
 675 eulysSerArgLeuGluGluValSerGluGluGluValArgGlu 691  
 134188 TTGAGCAAGATTTAGCAAGCAATCAGAGAACTCGAAGAAATCACT 134139  
 692 GluLeuSerSerLeuAsnGluPheLeuGluValGluValGluVal 708  
 134138 GACTTACATGTAATAATGACACTGCAAAAGAAATTTGCAAAAGTTTA 134089  
 708 sLeuAspAsnLeuLeuLeuPheGluGluValAsnAsnLeuThrGlu 725  
 134088 TTGTGATATAGCTCTGTAAGGACAGAGCAATCTTAACATTTGAA 134039  
 725 eLysAsn.....ValProLeuValValSerGluGluMetLysLysSer 739  
 134038 TGAATAATCATTAATGCTCTTAAAGTAAAGTAAAGCAATGAAAGCTCA 133989  
 740 HAspValLeuValAspAspLeuAsnLysLysLeuSerAspValThr 756  
 133988 CATGATGCAATTTGATGATGATGATGATGATGATGATGATGATGAT 133939  
 756 sLysArgThrGluLysLysLeuGluLeuGluValLeuMetGluAsn 773  
 133938 AAATATTACGAAAGAAAGAACTTGCAAAATGCAAAATTCCTACTGCAAAATG 133889  
 773 LeuSerLeuSerLysAsnValSerArgLeuGluValPheLeuPro 789  
 133888 ACGCTTATGATAGATGATGATGATGATGATGATGATGATGATGAT 133839  
 790 GluArgGluGluLysGluMetLeuLeuValSerAsnLeuLeuGlu 806  
 133838 GAGAAACATGAAAGAAAGATTAATGCTGTAATTCATTAATTCGAACT 133789  
 806 uLysLysGluLeuSerGluLeuAsnLysLysCysGluValProGlu 823  
 133788 TGAAGAAAGCTGTGTGATGATGATGATGATGATGATGATGATGAT 133739  
 823 yAlLeuValSerLeuMetSerGluAsnAsnAspLeuLysLysThrMetSer 839  
 133738 AAATTAACGCTTCACATCTGAAACACATCTGAAAGAAATGATGAGCT 133689  
 840 HLeuGluValProValLysThrHisGluGluLysLysThrLeuLeu 856  
 133688 ATCTGATGTATGTCAGATTTAAACGCTTAAACGCTTAAACGCTTAA 133639  
 856 rSerThrLeuAspLysThrAsnArgGluLeuValAspValLysLys 873  
 133638 TGAACGCTTGAAGCAAACTTACAGACATTAATTAATGATGCAAGAAAT 133589  
 873 yGluAspLeuAsnGluPheValLysLysAspGluValProGlu 889  
 133588 TTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 133539  
 890 LeuLysArgAsnLeuGluValAsnArgGluValLysLysLysLys 906  
 133538 TTAAGAAAGAAAGCTGCAAAAGCTGCAAGCAATTAAGCTGACAT 133489

906 sSerrLeuArgGluHisGluGluLysMetSerGlyLeuValGlySerMet 923  
 133488 CAGCTTGACAGACACAGAGCAAAAGATGAGCTGCTTAAGTCAGAGCATGA 133439  
 923 yLysValGluAspAsnSerValGluLeuValLysValLysLysSer 939  
 133438 GAAAGGTGAGAGATGATATGCTGATGATGATGATGATGATGATGATG 133389  
 940 GlnGluGluValLeuValThrLeuLeuGluGluValAlaLeuValArg 956  
 133388 CAAAGAGATTTGTCACATGTCATGCAAAATTAAGCCCAAGAAAGCA 133339  
 956 uLeuAspThrThrLeuGluCysLysLysLeuLysValAlaProIle 973  
 133338 GTCGACACACATTCACAGATGCTTAAGATGATGATGATGATGATGAT 133289  
 973 eTetGluGluLysGluValArgLysPheValAlaThrGluValArg 989  
 133288 CTTTGAAGATGTCGAGCAAAATTTAAACCAAGCAAGAAATCAATTA 133239  
 990 GluGluLeuSerGluGluThrGluLysValArgThrSerGluGluVal 1006  
 133238 GACCGATTAACAGACACACAAAGTATGCTGATGATGATGATGATGAT 133189  
 1006 sLysLysCysLysGluGluValAspLysLysLys..... 1017  
 133188 CAAAGAAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 133139  
 1017 ..... 1017  
 133138 CGAAATCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 133089  
 1017 ..... 1017  
 133088 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 133039  
 1017 ..... 1017  
 133038 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 132989  
 1018 ..... 1018  
 132988 AAAGCCGCTTCGAGGACCCCAAGCTTTCGCTGATGATGATGATGATGAT 132939  
 1029 AspLysAsnVal.....HisLeuGluAsnSerGlyCys..... 1039  
 132938 CCCAGCCGCTTCCCAATTAAGAGAGAGCTTTTCCCTTCGCAAGATTT 132889  
 1040 ..... 1040  
 132888 GAGCTAAGCAACCTCTCTTGAGGAGCTCATGATGAAAGGACGCTTA 132839  
 1045 eArgLysThrGluGluLeuValAsnArgLysLysAspLeuLeuVal 1061  
 132838 GCGGAAACAGAGAGCTTAACAAACAGCTTAACAAACAGCTTAACAAAC 132789  
 1062 TyrThrGluAlaLysLysGluLysLysGluValGluValAsnAla 1077  
 132788 TACACCGAAGTAAAGATGTAAGAAAGCAAGCTGTAAGAAAGAAATTC 132739  
 1078 LysGlnThrSerGluLeuAlaLeuGluThrLeuLeuGluValGluVal 1094  
 132738 AAACAGACTTCGAGATCTGATGATGATGATGATGATGATGATGATGAT 132689  
 132688 TGTTCATTTGCAACAGCTTGAAGGCTGCAAAATTCCTTAATGACCAAA 132639  
 1094 sValProLeuGluGluValGluSerLeuLysSerLeuSerGluVal 1111  
 1111 LeuLysArgGluGluLysLysThrLysGluValArgValCysValLys 1127  
 132638 TTGAAGATTAAGAAAGCAAGCTGATGATGATGATGATGATGATGATGAT 132589













[illegible]

|            |        |        |        |        |
|------------|--------|--------|--------|--------|
| BASE COUNT | 1594 a | 1023 c | 1080 g | 1228 t |
| ORIGIN     |        |        |        |        |

```
alignment_scores:
  Quality: 1248.50      length: 1421
  Ratio: 1.786          Gaps: 33
  Percent Similarity: 49.191  Percent Identity: 26.038
```

```
alignment_block:
US-09-750-590-2 x AF155135 ..
```

Align seg 1/1 to: AF155135 from: 1 to: 4925

[illegible]

754 GATCTTGTGTAACAAAGCTTCATTATTCACAAGCTCAGAAAAAGC 803  
233 TAspIleLeuThrLeu..... 239  
804 AGCAATTTCAAGAAGCTCATATCATCAAATAATCTCAGATCTGCATTAA 853  
240 ..... LysThrAlaSer 243  
854 AACCCCGAACAAAACCAAGACAGATGACCAACATCTGCTAAATAAGCTA 903  
244 GluHisSerAsnLysGlyArgIleLeuTrpLysGlyProSerLeuI 260  
904 GAAGGAAGGGAAAGCTCCAAAACACGCCAAAGCTCACACCCTCTCATACG 953  
260 nGlnArgAsnLeuSerGlnMetLeuAspValLaaMethLysSerPheG 277  
954 TCACACCCACTGATGATGATGCTCTCTCCCAAGATCAATTAAGCTTGACATC 1003  
277 IAsrGlyHisGlnAsnLcGlnAsnPheGluLileLysaGlnLAspLeu 293  
1004 CACTATCGGGCAAG... GAATGGTATTTTTTGTGCAACCAACCTCTC 1047  
294 LysGluArgLeuArgLysIleGlnGlnGlnGlnIleArgIleLeuLeuAsp 310  
1048 AAGCGTAGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1097  
310 ATGAAAGGlyLeuGlnLeuGlnLeuAsnGluLysValIleValAlaAspA 327  
1098 TTCTACAGCT..... GCTGATA 1114  
327 AspLeu..... GluSerGlyGlyGlyLeuLysSerLeu 338  
1115 GCTTATGTCGATATTAAGTCTGACAGTCTGACCAACATCTTCTCTCTCTA 1164  
339 LeuAlaIleLysGlyArgLysGln..... HisGluGlySerLeuArgTh 352  
1165 TTGCACGCAAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1214  
352 rLIleGluAlaLeuLysSerArgPheHisLysTrpPheGlnSerAspHisLeuG 369  
1215 TAAATTCAGAGCAAAATCC.... MAGAGAGCGCAAGACAGCATTAAGCT 1261  
369 LysGlySerLeuSerPheArgLysGlyLysaGlnMet..... LeuLeuLys 382  
1262 TTGACATCAACACATTCACGCCAAAGTACTGTGGCCATCTCCCTCGGAAA 1311  
383 GlnGlyGlnIleMetTrpMetThrAspSerGlnCysHisSerThrGlyMetPr 399  
1312 CCGTGATCAAACTCTCCCCCGAGCTCCAAA... TCATCTGCATCTGTCTT 1358  
399 oVALHISMeIGlnSerArgSerMetLeuArgFroLeuGluLeuAlaLeuP 416  
1359 AAATCAATCTTTACGTAAATTC..... 1380  
416 rGMSGlnAlaSerTrpSerGluAsnGlnIleLeuLysGlyGluLeuGlu 432  
1380 ..... 1380  
433 AlalMetArgThrPheCysAspSerAlaLysGlnAspArgLeuLysLeuG 449  
1380 ..... 1380  
449 pAsnGluLeuAlaHisLysValAlaGluCysLysAlaLeuAlaLeuGluC 466  
1380 ..... 1380  
466 YsgLysValLysGlnAspSerArgGlnIleLysGlnAlaGluAsp 482  
1381 ..... ACACAGCAAGATGATGTCAGGATTCAGACCTGCAAG 1419  
483 AlaleuLysAspValGlnLysArgMetCysGluSerGlnCysLysValLys 499  
1420 ATTTCCTGATCATCTACAGAGAGAGAAAGAGAGCTCTGAAGCAGAGGAAA 1469

```

499 scdMetGlnThrHisPheLeuAlaLeuLysGlnHisLeuThrSerAspA 516
|||||
1470 ACAGCTACAG..... 1479
516 IaAlaThrGlyAsnHisArgLeuMetGlnGluLeuLysAspGlnLeuLys 532
1479 ..... 1479
533 AspMetLysValLysTyrGlnLysAlaSerAlaValGlyGlyLysLeuA 549
|||||
1480 .....GTCAACTCCATCCGCAAGCGCACTGGATGCTTA 1520
549 GAsnGlnLeuLysGlnAsnGlnMetLeuValGlnGluPheLysArgAsp 566
1521 CAC..... 1524
566 LucGlyLeuMetGlnGlnLysAsnLysArgLeuGlnLysGluLeuSerMet 582
1524 ..... 1524
583 CysGlnLeuGlnLysArgGlyValArgLysLeuThrGlnMetGlnG 599
|||||
1525 .....ACTGAGATTTCAG 1538
599 yGlnLeuLysAspLeuSerAlaLysLeuAlaLeuSerLThrGlnGlu 616
|||||
1539 GAACAGCTGCACTCAAG..... 1557
616 ySphGlnLysAsnMetLysSerLeuLeuSerAsnGlnLysAsnGlnLysAla 632
1557 ..... 1557
633 LysLysLeuAlaSerValGlnLysGlnLysArgLysSerLeuAsnGlnL 649
|||||
1558 CAGAACTTAAGAAAGCTCAAGCAAAATACGAGGAGCTATCAAAAGAG 1607
649 PArgProLeuLysArgGlnLeuGlnLysAsnLeuLysAlaLysLeuAlaGln 666
1608 CTTAGCTGTCCAGAACAGATG.....AACTGGCTCTT 1641
666 LysValLysProGlnGlnHisGlnGlnLeuLysSerArgLeuGlnLys 682
|||||
1642 ..GTCTACCTGAACAGATGATTAATATTCTACATTCCAGAG..... 1683
683 SerGlnLysLeuGlnLysArgGlnThrGlnLeuThrSerLysAsnGln 699
1684 .....CTAGAGGTCAAGAA..... 1698
699 rLeuGlnLysGlnLeuGlnLysValCysLeuAspAsnLysLeuLeuThrG 716
1699 .....G 1699
716 InGlnValAsnAsnLeuThrThrGlnMetCysAsnValProLeuLysVal 732
|||||
1700 AGCAATTAATATGCTTAACAGCATCTCCAGAAATCCA.....TTAAAGAA 1746
733 SerGlnGlnMetLysLysSerHisAspValLevalAspAspLeuAsnLys 749
|||||
1747 AGTCAAGAAATTA..... 1761
749 sLysLeuSerAspValThrHisLysTyrThrGlnLysLysLeuGlnMetG 766
|||||
1762 .....GAGAAATGAGAGAGTTAG 1780
766 LysLysLeuMetGlnAsnAlaSerLeuSerLysAsnValSerArgLeu 782
|||||
1781 AGCAAAAGCTGTAGAG.....AGCAG 1803
783 GluThrValPheLThrProGlnLysGlnLysGlnMetAlaLe 799
1804 AAGCTACAGATTAAGCA..... 1824
799 uLysSerAsnLThrGlnLeuLysLysGlnLeuSerGlnLeuAsnLysAl 816
1824 ..... 1824
816 yScyGlnGlnLysAspGlnGlnLysLThrSerLeuMetSerGlnAsnAsn 832
1824 ..... 1824
833 AspLeuLysLThrMetSerAsnGlnLysValProValLysThrHisG 849
|||||
1835 .....CTGTGAGAGAGATCA 1841
849 uGlnLysThrAlaLeuSerSerThrLeuAspLysThrAsnArgGln 866
|||||
1842 GCAATGAAAGTTCAATGCTCTGTATGCAATGCAATGAGAG..... 1890
866 euValAspValLysLysCysGlnAspLLeuAsnGlnLysPheValLys 882
1890 ..... 1890
883 IleLysAspGlnAsnGlnLLeuLysArgAsnGlnLysAsnThrGlnAs 899
1890 ..... 1890
899 nGlnValLysAlaGlnLysLThrSerLeuArgGlnLysGlnLysMetS 916
1890 ..... 1890
916 eGlyLeuLysArgLysSerMetLysLysValGlnAspAsnSerAlaGln 932
1891 .....AAAGCTATTTCG 1902
933 LeuAlaLysTyrLysLysSerGlnGlnGlnLysLevalThrLeuHisGln 949
|||||
1903 TTGAAGAAATCCAGAAAGCCGCAAGAAATATGAAATTAAGAC..... 1950
949 uLeAlaLysGlnLysArgGlnLysAspThrLThrGlnLysLysL 966
1950 ..... 1950
966 euLysTyrAlaProLThrLeuSerLeuGlnGlnLysGlnLysArgPheLys 982
1950 ..... 1950
983 AlaThrGlnLysGlnLeuLysGlnGlnLeuSerGlnLThrGlnLysTyr 999
1991 .....ACATTAAGATGATCAAGCAAGAA..... 1977
999 rAsnThrSerGlnGlnLysAlaLysLysCysLysGlnGlnLysAsnLysL 1016
1978 .....GCCAGTATGAGCT..... 1992
1016 euLysLysGlnLLeuThrLThrLeuGlnLysAspLeuLysAspVal 1032
1993 ..... 2007
1033 HisLysLysAsnSerTyrGlnGlnLysAlaLeuSerArgLysThrGln 1049
2008 .....GCCATGAATGAGATGATGATG 2027
1049 uGlnLeuAsnArgGlnLysLysAspLeuLeuGlnLysTyrThrGlnAlaL 1066
|||||
2028 TGAATCAATTAAGCAAGATGAGCAAGCTGTCCAAAGAAAGCC 2077
1066 yLysGlnLysGlnLysLeuValGlnGlnLysAsnAlaLysGlnLThrSerGln 1082
2078 AGCTTACCTGTGAG.....GATTACAGAGAGAGAAATCTCATAG 2118
1083 rLeuAlaLysGlnThrLeuLeuGlnLysGlnLysValProLeuGlnG 1099
2119 GATGTCAAGCTCAATGATATGCTAATTAAGCAAGCAATGAGAGAGAG 2168
1099 nValGlnSerLeuLysLysSerLeuSerGlnLThrLThrGlnLysLeuLys 1116

```



```

33 pValGluIyValSerSerIleuLeuAlaIySylGlyValAsnProGlyL 50
121 TGCTGAGAAAGCTGCTCCTGCTGCGCAAAAGAGGCGCCAGCGCCACA 170
50 yLeuAspValGluGlyArgSerAlaPheHisValAlaIaSerIyGly 66
121 AGCATGACAGAGGAGCAAGCCGCTCCCTCCCTCCGCGAGGGA 220
67 AsnLeuGluCysLeuAsnAlaIleuLeuIleuHisGlyValAspIleuPhe 83
221 CAGTGGATGCTCCAGAGCTCATGGTGAACAGCGGTGGATGTGATCTCC 270
83 rSerAspThrAlaGlyArgAsnAlaLeuHisLeuAlaAlaIySylGly 100
271 GCGAAATCTTCCTGCGACAGAGTCTTCGATCGACAGCCAGAAATGGCA 320
100 lAlaLeuGluCysLeuGlnIyLeuLeuLeuIleuTyrAsnProPhuGluHis 116
321 ACCGAGAGCATCAGAGACTTCTTCAGTACAAAGAGCCAGCGAGAAC 370
117 ValAspLeuGlnIyArgThrAlaLeuHisAspAlaAlaIaMetAlaAsp 133
371 ATTGACACACCTGGAGAAACAGCTTCGATTCGATTCAGAGGACAGGAG 420
133 rProSerSerIleGlnIleuLeuCysAspHisGlyAlaIaSerValAspAla 150
421 CCTTCAGCGCTACAGCTCCTCGGACACAGAGCCCATTAACCTTA 470
150 yAspValAspGlyArgThrProLeuValLeuAlaIaIleuIleuCysArg 166
471 AAGACTTGGATGGAAACATCCCGCTCGCTGCTGCTGACAGAGAGCGCAC 520
167 ProThrIleGlyGlnIleuLeuIleuAspArgGlyAlaAspIleuAsnSer 183
521 AGTGAGCGCTCCACCTTCCTCGTGGATCCGCGCGCATGTCAATTCGAG 570
183 gAspLyGlnAsnArgThrAlaLeuMetLeuGluCysGluTyrGlyAla 200
571 AGRCAAAATGGAGAACACTGCTGCTGATGGCTGTGGTGAACCGGAGCT 620
200 yAspAlaIyAlaGlyValLeuIleuIleuAsnGlyValAspValIleuLeu 216
621 CGAACACTGTGAGCGCTTAATTAAGAGGACGGAGACCGACTGTGC 670
217 AspAlaIeugIyHisAspSerSerTyrTyrAlaGlyIleuIAspAsnIle 233
671 GATTCCTCGGACAAATGGCTTACATTAATTCACAAACCTCCGAAATTC 720
233 uAspIleuIleuThrLeuLeu 239
721 AGSAAATTCAAACCTTCATATTATCAAAAATGCTGAGGATGCTATTAA 770
240 .....LysThrAlaIaSer 243
771 AGACTCCAAACAAACCAAGCAGCATGACCAAGTCTCTCAAAAATTAAGCTA 820
244 GluAsnSerAsnIyGlyArgGluLeuThrIyValGlyProSerLeuG 260
821 GAGGAAATGGAACTCCAAAGAGCGCAAGCTCCAGCCACTCTCTATG 870
260 nGlnIArgAsnIleuSerGlnMetLeuAspGlyValAsnThrIySerAsnG 277
871 TCTTACCCAGTGAATGATGTGCTCTCCCGCATCAATTAATCATATCAC 920
277 lAspArgIleuHisGlnAsnIleuGlnAspLeuGluIleuGluAsnGlu 293
921 CACCTTCAGAGAGG.....GAGTCAGTATTTTGTCTGAGAGCCCTTT 964
294 LysGluArgLeuArgIleuGlnIyGluGlnIyAlaGluIleuLeuAspIy 310
965 AAGGCTGAGATCAGCTCATACAGAAACAAAGAGCGCTGTACGACAG 1014

```

```

310 sValAsnGlyLeuGlnIleuGlnIleuAsnGluGlyValIleuValAlaAsp 327
1015 TACTGCAAGT.....GluSerGluIyGlyValLeuIySerLeu 338
327 sPLeu..... 338
1032 GCTTGTGGATATATGATTCTTGAGAGCTGACAGCAACAGATCTTGTCCTG 1081
339 LeuAlaAlaIyGluIyGln.....HisGluGluSerLeuArg... 351
352 rThrIleGlnIleuAlaGluSerArgPheIyGlyThrPheAlaIaSerAspHisL 368
1132 TAAATTGCGGCCAATTCGCCAGAGACAGCAAGAGCGAAAGCCGACTTA 1181
368 euGlySerIlySerHisPheAlaArgIyGlnAspMetLeuLeuIyGlnGly 384
1182 GCTTCAGCTTCCTTCATTCAGCAACAGCTGACCTGCGCCCTCCAGGCG 1231
385 GlnMetIyMetThrAspSerGluCysThrIleuThrIleuMetProValHis 401
1232 AAGCTAGAGCATCCCTCTCTCCAGATCGCAAAATCATCTCCACCCCTTGA 1281
401 sMetGlnSerArgSerMetLeuArgProLeuGluLeuAlaIleuAsnG 418
1282 GCAT.....CCGCGCG 1292
418 lAlaIaSerTyrSerGluAsnGluIleuLeuIySylGluLeuGluAlaIaMet 434
1293 GAGCGTCCACGCTGACATGATGTCATA.....CCGCGCG 1321
435 ArgThrPheCysAspSerAlaIyGlnAspArgIleuIySylGlnAsnG 451
1321 ..... 1321
451 uLeuAlaHisIyValAlaIleuGluCysArgIyAlaLeuAlaIeugIyGlu 468
1321 ..... 1321
468 rValIyGlnAspSerAspGluGlnIleuGluIleuGluAspAlaIe 484
1322 .....ATCCGCGACCTCCCAAGATCTCCG 1345
485 LysAspValGluIyArgMetIyGluSerGluGlyIyValIySylGlu 501
1346 CATGATCTCCAAAGATGGAGAGAGCTGAGAGAGCGCAAGAAAGCTCGAGAG 1395
501 tGlnIThrHisPheLeuAlaIleuIySylGlnIleuIleuHisSerAspAlaIa 518
1396 CCGAGCAGCA.....CTCCAGTCAACAAAGAGCGAGCAGCGTTCT 1436
518 hGluIyAsnHisArgIleuMetGlu.....GluLeuIyAspGlnIleu 531
1437 TGAACAAACAGAAATCTCCGAGAAAGCTCTGATACCTGATGAGAAAGCTC 1486
532 LysAspMetIyValIySylGlyGluIyValSerAlaIleuValIyGlySyl 548
1487 AAGGAAACCGAGAGAGTACGAGAGAGCGCAAGAAAGAGCTCGAGAGCT 1536
548 uArgAsnGlnIleuGluIleuGlnMetLeuValGluGluPheIyArg 565
1537 GCGAAACAGAAATGAGCTG..... 1555
565 sProlGlyIySylLeuMetGluGluIyGlnIySylArgIleuGlnIyGluLeuSer 581
1556 .....GGTCTCTCTCCCGAGAGAGC..... 1576
582 MetCysGluLeuGluIyArgGluIyArgGlyIySylLeuThrGluMetG 598
1576 ..... 1576
598 uCylGlnIleuIyAspIleuSerAlaIyLeuAlaIleuSerIleuProIleu 615

```

```

1577 .....|111:
615 LulyspHeGluAmMetLysSerLeuSerAsnGluLeuAmGluLys 631
      :|:
1581 ATGGCTACTGCGATCTCCGGCAG..... 1603
632 AlaLysLysLeuLeuAspValGluArgGluTyrGluArgSerLeuAsnGlu 648
1603 ..... 1603
648 uThrArgProLeuLysArgGluLeuGluAsnLeuLysAlaLysLeuAsnGlu 665
      |||
1604 .....GCCCGACGTCAGCAAGACATCGACACACTGAAG..... 1636
665 LnhLysValLysProGluGluLnhLysGluGluLeuLysSerArgLeuGluGln 681
1636 ..... 1636
682 LysSerGlyGluLeuGlyLysArgGluLeuThrGluLeuThrSerLysAsnGlu 698
1637 .....CA 1638
698 nThrLeuGlnLysGluLeuGlnLysValCysLeuAspAsnLysLeuLeuT 715
      |||||
1639 AGACCTCGAAGAAAGCGTGGCA..... 1660
715 hrcGlnGlnValAsnAsnLeuThrThrThrGluMetLysAsnValProLeuLys 731
1660 ..... 1660
732 ValSerGluGluMetLysLysSerHisAspValIleValAspAspLeuAs 748
1660 ..... 1660
748 nLysLysLeuSerAspValThrHisLysTyrThrGluLysLysLeuGluLm 765
1660 ..... 1660
765 eGluLysLeuLeuMetGluAsnAlaSerLeuSerLysAsnValSerArg 781
      |||||
1661 .....GAAGTCGACAGGACAGGAGCGGCGCGCGAG 1693
782 LeuGluThrValPheLeuProProGluArgHisGluLysGluMetAla 798
      |||||
1694 TTGACACTAGCTT..... 1708
798 aLeuLysSerAsnIleThrGluLeuLysLysGlnLeuSerGluLeuAsn 815
1709 .....CCGAGAAAGACAGCGCGCAAGCGCACTA 1736
815 yelLysCysGluLysAspGlnGluLysLleLysSerLeuMetSerGluAsn 831
1737 ACCCA..... 1741
832 AsnAspLeuLysLysThrMetSerHisGlnTyrValProValLysThrH 848
1742 .....CCGGCCGACGCGC 1755
848 sGluGluLysLysThrAlaLeuSerSerThrLeuAspLysThrAsnArg 865
1756 TGAGGACGTGAGAACTTCCTGCTGCTGTGTGCAAAATGATTAAG 1805
865 LuLeuValAspValLysLysCysGluLysAspHisLeuGlnLysPheVal 881
1806 AG..... 1807
882 LysIleLysAspGluAsnGluIleLeuLysArgAsnLeuGluAsnThrG 898
1807 ..... 1807
898 nAsnGlnValLysAlaGluTyrIleSerLeuArgGlnIleGluLysAm 915

```

```

1807 ..... 1807
915 eLSerGlyLeuArgLysSerMetLysValGluAsnAspSerAlaGlu 931
1808 .....ACGCATT 1816
932 IleLeuAlaLysTyrLysLysSerGlnGluIleValThrLeuLys 948
1817 TCTTTGAAATATTCAGACGCCCAAGAAATCATCAAACTGAAAGCA 1866
948 uGluIleAlaIleGlnLysArgGluLeuAspThrIleGlnLysVal 965
1867 ..... 1867
965 yLeuLysTyrAlaProIleIleSerLeuGluGluCysGluArgLysPhe 981
1867 ..... 1867
982 LysAlaThrGlnLysGluLeuLysGlnLeuSerGlnIleThrGlnLys 998
1868 .....ACGCTAAAGATCGATCGACAGAGAGCCG..... 1900
998 sTyrAsnThrSerGluGluGlnAlaLysLysCysLysGlnLysAsnAsp 1015
1900 ..... 1900
1015 yLeuLysLysGluIleLeuThrLeuGlnLysAspLysAspLysAsn 1031
1900 ..... 1900
1032 ValHisIleGluAsnSerTyrGluThrGluArgAlaLeuSerArgLysThr 1048
1901 .....GACGACGTGGGATATGATGAAGGCGCATGAACAGATCAT 1941
1048 rGluGluLeuAsnAspArgGlnLeuLysAspLeuLeuGlnLysTyrThrGlu 1065
1942 TGATGACTCAATTAAACAGGTGACGACGTCTCACTGACGCAAG 1991
1065 lAlaLysGluLysGluLysLeuValGluGluAsnAlaLysGlnThrSer 1081
1992 CCCAGGCGAG.....CTGAGAGATTACAGAGAGAGAAATCT 2029
1082 GlnIleLeuAlaIleAlaThrLeuLeuGlnGlnLysGlnHisValProLeu 1098
2030 CTGGAGATCGCCCGGACGTACATCAACAACTGAAACAGCGCAT 2079
1098 uGlnValGlnSerLeuLysLysSerLeuSerGlyThrIleGluThrLeu 1115
2080 GCACGTGTCAACCTCTCCAGGCGCAAGTGC.....GAAAGAGCC 2120
1115 yGluGluLeuLysThrLysGlnArgGlyTyrGluLysGluGlnIleThr 1131
2121 TGTGCGAGTGAATGCTCAG.....TATTCGAAAGGTTTGAAAGG 2161
1132 ValThrGlnLeuArgIleMetLeuGluLysGlnLysAsnSerSerValPr 1148
2162 TTGATCACTGCTAAACACTGTGGATGACACCAAGCAACTGCTATC 2211
1148 oLeuAlaGlnLysLysGlnValLysGluAlaPheGluLysGluVal 1165
2212 CATCAAGAGCACTTTCGAAGT.....ATTAACCA 2240
1165 IleIleLysAlaSerLeuArgLysGluGluGluSerGln..... 1178
2241 CGCTCGGCAACCGCAAGAGATGAGAAAGAAATGAACGCGCTGACA 2290
1179 .....AnLysThrGluGluLysLysSerGlnGlnIleSerGlu 1191
2291 GGCACCTTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
1191 eGlnAsnThrLysGlnAlaLeuLysLysLeuLysThrArgGluValAla 1208
2341 CGCCGAG 2378

```



```

613 CTTTTAGCTGTACAGCTCTCTGTGACAGCAAGACCCCAATTAACCTTA 662
150 ysaapvalaapglvargthprolevalaleuallthglnlcyarg 166
663 AAGCTTCTGGTGAACATCCCTCTCTCTGTGTCTGTGAAACCGGACAC 712
167 Prothrllecysglnleuenuileasparaglylaasparileuaspar 183
713 AGTGAGCGCTGCACCTTCTCTGTGATCGACGCGCAAGTGCATTAATCCAG 762
183 gaaplyaglnaasargthralaleuallcysgllytrglnlcyarg 200
763 AAGCAAAATGGAAACATCGCTCATCTGCTCTGTGAAACCGGACGCT 812
200 ysaapalvalglvalleuileysanndlyalaasparalytrleu 216
813 CGAAGCTGTGACGCCCTTAATTAAAGGCGCACTGACATCCCTGTCTC 862
217 Aapalaleucllylaasparsetrtytralaarglileciysaapala 232
863 CATCTCTGGACACAAATCCCTACATTTATCCAAATCTGTGGAAATGCG 912
233 uaaplleuethrileuenu... 239
913 AGGAATTGAAACCTCTATTATCAAAATATCTGTGAGATCTGAATTAA 962
240 ... 1012
963 AGACTGCACAAACCAAAAGCAGCATGACCAACATCTCTAAATTAAGTCA 1012
244 Glnuanserhnllysglyarglyleuarglyleuarglylproserleu 260
1013 GAAAGCACTGAAATCCAAAAAAGGCGCAACGTCACCATCTCTATCAG 1062
260 nglatgaaaleuaserfclmetleuapglvalaasparhlyseraag 277
1063 TCTACACCTGTGATGATGTCTCTCTCCCGCATCAATTAACATCTACAC 1112
277 lnatrglnlueglnaasparleuileuileuamnglnuileuamnglnuileu 293
1113 CACTTTCAGGAAG...GACTCAGTATTTTTTGTCTGAAGCGCCCTTT 1156
294 lysgluatrgleuarglylglnglnuileuamnglnuileuamnglnu 310
1157 AAGCTGAGATCAAGCTCATACGAAACCAAGACAGCTGAGACAG 1206
310 aylaaleuileuileuileuileuamnglnuileuamnglnuileuamnglnu 327
1207 TACTCGACGCT...GCTGACA 1223
327 spieu...Glnsercllysglnlyleuileuileuileuileuileu 338
1224 GCTTCTTGGAATATAAGTTCGAAGCTGACCAAAAGATCTCTTGTCTCG 1273
339 leuallalelysglnlyleuileu...HlsGlnllyserleuarg... 351
1274 TTGCAGACAAAGCTCGCTCTCTTACCTTCAACATTAAGAGATCTGAGA 1323
352 'Thrlleuileuileuileuileuileuileuileuileuileuileuileu 368
1324 TAAATTCAGACCAATGCCCAACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1373
368 eucllysercllyserhlnphearglyleuamnglnuileuamnglnuileu 384
1374 GCTTTCAGCTTTCATTCGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1423
385 Glnmetrlymetthrasparfclucytrhserthrlglnmetproval 401
1424 AAGACTAGAGACATCCCTCTCTGATCGACCAATCAATCCCAACCTTCA 1473
401 smedclnserargsermetleuargproleuamnglnuileuileuileuileu 418

```

```

1474 CCAT...CCGGCG 1484
418 lvalaserfcllyuamnglnuileuileuileuileuileuileuileuileu 434
1485 GGACGCTGCACCACTGACCAATGATGTGATA... 1513
435 Argthphecysaasparalalyglnasparargleuileuileuileuileu 451
1513 ... 1513
451 uleuallalelyvalaleuileuileuileuileuileuileuileuileu 468
1513 ... 1513
468 rgallylsglnaasparaspgluileuileuileuileuileuileuileu 484
1514 ...ATCGGACGCTGCAAATTTCTCG 1537
485 lysaapvalaaleuileuileuileuileuileuileuileuileuileuileu 501
1538 CATGATCTCGAAGAGATTGCAAGATCTGACGACGAGCAAGCAAGCAAGCT 1587
501 tclnthraspaleuileuileuileuileuileuileuileuileuileuileu 518
1588 CAGAGACAA...CTCAATGTCAAGAGCAAGCAAGCAAGCTTCTCT 1628
518 hcllyasnhlaleuileuileuileuileuileuileuileuileuileuileu 531
1629 TGCAACACAGAGATCTCGAAGATGCTGTGATCTGATCTGAGACAGCTC 1678
532 lysaasparleuileuileuileuileuileuileuileuileuileuileuileu 548
1679 AAGCAACCCAGAGCAAGATGCAAGAGCAAGCAAGCAAGCAAGCAAGCT 1728
548 uarghnglnlileuileuileuileuileuileuileuileuileuileuileu 565
1729 GCAGAAACAGATGCAAGCTG... 1747
565 spcllyllyleuileuileuileuileuileuileuileuileuileuileuileu 581
1748 ...GATCTCTCTCGAGACAG... 1768
582 Metcysglnuileuileuileuileuileuileuileuileuileuileuileu 598
1768 ... 1768
598 ucllylneleuileuileuileuileuileuileuileuileuileuileuileu 615
1769 ...CCGG 1772
615 lulyasphesglnuileuileuileuileuileuileuileuileuileuileu 631
1773 ATGCTACTGCTGATCTGAGAC... 1795
632 Alalylyleuileuileuileuileuileuileuileuileuileuileuileu 648
1795 ... 1795
648 uthrtrproleuileuileuileuileuileuileuileuileuileuileuileu 665
1796 ...CGCGCACTGACCAAGATGCAAGCAAGCAAGCAAGCAAGCAAGCA 1828
665 lnhllyvallyprogluileuileuileuileuileuileuileuileuileu 681
1828 ... 1828
682 lyseserllylueuileuileuileuileuileuileuileuileuileuileu 698
1829 ...CA 1830
698 nthleuileuileuileuileuileuileuileuileuileuileuileuileu 715
1831 AGACTGTGAGAAAGCGGTGAA... 1832

```



```

715 hrGInGInValAsnAsnLeuThrThrGluMetLysAsnValProLeuLys 731
1852 .....
732 ValSerGluMetLysLysSerHisAspValIleValAspAsnLeuAs 748
1852 .....
748 nLysLysLeuSerAspValThrHisLysTrpThrClnLysLysLeuGlu 765
1852 .....
1852 .....
765 etGluLysLeuMetGluAsnAlaSerLeuSerLysAsnValSerArg 781
1853 .....
1853 .....
782 LeuGluThrValPheIleProProGluThrHisGluLysGluMetIleAla 798
1886 .....
1886 .....
798 aLeuLysSerAsnIleThrGluLeuLysGlnLeuSerGluLysAsnAla 815
1901 .....
1901 .....
815 LysLysGluGluGluLysAsnGlnLysLysLysSerLeuMetSerGluAsn 831
1929 .....
1929 .....
832 AsnAspLeuLysLysThrMetSerHisGlnTrpValProValLysThrHI 848
1934 .....
1934 .....
848 sGluGluLysThrAlaLeuSerSerThrLeuAspAlaSerLysAsnArg 865
1948 .....
1948 .....
865 LLeuValAspValLysLysLysCysGlnAspIleAsnGlnLysVal 881
1998 .....
1998 .....
882 LysIleLysAspGluAsnGlnLysLeuLysArgAsnLeuGluAsnThrGlu 898
1999 .....
1999 .....
898 AsnGlnValLysAlaGluTrpLleSerLeuArgLysIleGlnLysLys 915
1999 .....
1999 .....
915 etSerGluLeuArgLysSerMetLysLysValGlnAspAsnSerAlaGlu 931
2000 .....
2000 .....
932 IleLeuAlaLysTrpLysLysSerGlnGlnLysIleValIleThrLeuHisGlu 948
2009 .....
2009 .....
948 uGluIleAlaIleGlnLysArgLysLeuAspThrIleGlnLysCysLysIle 965
2059 .....
2059 .....
965 LysLeuLysTrpAlaProIleIleSerLeuGlnGlnCysGlnArgPhe 981
2059 .....
2059 .....
982 LysAlaThrGluLysGlnLysGlnLysGlnLysSerGlnGlnIleThrGlnLys 998
2060 .....
2060 .....
998 sTrpAsnThrSerGlnGlnLysAlaLysLysCysLysGlnLysAsnAspL 1015
2092 .....
2092 .....
1015 LysLeuLysLysGlnLysIleLeuThrLeuGlnLysAspLeuLysAspLysAsn 1031
2092 .....
2092 .....
1032 ValHisIleGlnAsnSerTrpClnLysArgAlaLeuSerArgLysThr 1048
2093 .....
2093 .....
1048 rGlnLysLeuAsnArgGlnLeuLysAspLeuLysGlnLysTrpThrLys 1065
2134 .....
2134 .....
1065 LAlaLysGlnLysGlnLysLysValGlnLysAsnAlaLysGlnIleThrSer 1081
2184 .....
2184 .....
1082 GlnIleLeuAlaIleAlaThrLeuGlnGlnLysGlnIleValProLeuGlu 1098
2222 .....
2222 .....
1098 uGlnLysGlnSerLeuLysLysSerLeuSerGluTrpThrIleGlnLysLys 1115
2272 .....
2272 .....
1115 LysGlnLysLysTrpThrLysGlnArgCysTrpGlnLysGlnLysIleThr 1131
2313 .....
2313 .....
1132 ValThrGlnLeuArgGlnMetLeuGlnAsnGlnLysAsnSerSerValPr 1148
2354 .....
2354 .....
1148 oLeuAlaGlnIleLeuGlnValLysGlnAlaPheGlnLysGlnValCysI 1165
2404 .....
2404 .....
1165 LLeuLysLysSerLeuArgGlnLysGlnLysGlnLysSerGln..... 1178
2433 .....
2433 .....
1179 .....
2483 .....
2483 .....
1191 eGlnAsnThrLysGlnAlaLeuLysLysLysGlnIleThrArgGlnValValAla 1208
1253 .....
1253 .....
1208 sPheLysTrpTrpLysAlaThrLysSerAspLeuGlnIleIleSer 1224
2571 .....
2571 .....
1225 AspLeuAsnGlnLysLeuAlaAsnLeuAsnArgLysTrpGlnGlnValArg 1241
2621 .....
2621 .....
1241 sGlnGlnValAlaLeuHisAlaLysLysLysGlnLeuSerAlaLysAspGln 1258
2671 .....
2671 .....
1258 LysGlnLeuLeuHisPheSerIle.....GlnGlnIleIleLysAsp 1271
2721 .....
2721 .....
1272 ArgGlnAsnArgProCysArgCysTrpGlnAlaGCGAAAGCGAGGAGTACTGCG 1288
2771 .....
2771 .....
1288 rArgIleGlnLysSerAlaLysGlnIleGlnAlaLysAspAsnLysIleTr 1305
2821 .....
2821 .....
1305 hrClnLeuLeuAspAspValGlnArgLeuLysGlnAlaLeuAsnGlnLysLeu 1321

```

|       |        |                            |
|-------|--------|----------------------------|
| 1     | 887:   | contig of 887 bp in length |
| 888   | 1780:  | contig of 993 bp in length |
| 1781  | 2645:  | contig of unknown length   |
| 2646  | 3525:  | contig of 865 bp in length |
| 3526  | 4401:  | contig of 880 bp in length |
| 4402  | 5290:  | contig of 876 bp in length |
| 5291  | 6155:  | contig of 889 bp in length |
| 6156  | 7151:  | contig of 865 bp in length |
| 7152  | 8036:  | contig of unknown length   |
| 8037  | 8948:  | contig of 885 bp in length |
| 8949  | 9836:  | contig of 912 bp in length |
| 9837  | 10705: | contig of unknown length   |
| 10706 | 11579: | contig of 869 bp in length |
| 11580 | 12459: | contig of 874 bp in length |
| 12460 | 13331: | contig of 880 bp in length |
| 13332 | 14190: | contig of unknown length   |
| 14191 | 15058: | contig of 862 bp in length |
| 15059 | 15923: | contig of 868 bp in length |
| 15924 | 16802: | contig of 885 bp in length |
| 16803 | 17680: | contig of unknown length   |
| 17681 | 18548: | contig of 878 bp in length |
| 18549 | 19413: | contig of unknown length   |
| 19414 | 20289: | contig of 868 bp in length |
| 20290 | 21156: | contig of 855 bp in length |
| 21157 | 22031: | contig of 876 bp in length |
| 22032 | 22916: | contig of unknown length   |
| 22917 | 23786: | contig of 883 bp in length |
| 23787 | 24655: | contig of 870 bp in length |
| 24656 | 25519: | contig of 876 bp in length |
| 25520 | 26383: | contig of 869 bp in length |
| 26384 | 27253: | contig of 864 bp in length |
| 27254 | 28119: | contig of unknown length   |
| 28120 | 28987: | contig of 866 bp in length |
| 28988 | 29861: | contig of 868 bp in length |
| 29862 | 30737: | contig of 874 bp in length |
| 30738 | 31601: | contig of unknown length   |
| 31602 | 32481: | contig of 864 bp in length |
|       |        | contig of 880 bp in length |

| Accession | Contig                            | Length | Gap | Percent Similarity | Percent Identity                  |
|-----------|-----------------------------------|--------|-----|--------------------|-----------------------------------|
| 32482     | 33347: contig of 866 bp in Length | 866    | 0   | 6357               | 6424: contig of 866 bp in Length  |
| 33348     | 34218: contig of 871 bp in Length | 871    | 0   | 64625              | 65487: gap of unknown length      |
| 34219     | 35081: contig of 863 bp in Length | 863    | 0   | 65488              | 66346: gap of unknown length      |
| 35082     | 35948: gap of unknown length      | 863    | 0   | 66347              | 67207: contig of 861 bp in Length |
| 35949     | 36814: contig of 866 bp in Length | 866    | 0   |                    |                                   |
| 36815     | 37687: contig of 873 bp in Length | 873    | 0   |                    |                                   |
| 37688     | 38583: contig of 896 bp in Length | 896    | 0   |                    |                                   |
| 38584     | 39452: gap of unknown length      | 896    | 0   |                    |                                   |
| 39453     | 40334: contig of 882 bp in Length | 882    | 0   |                    |                                   |
| 40335     | 41201: contig of 867 bp in Length | 867    | 0   |                    |                                   |
| 41202     | 42067: contig of 866 bp in Length | 866    | 0   |                    |                                   |
| 42068     | 42933: gap of unknown length      | 866    | 0   |                    |                                   |
| 42934     | 43801: contig of 868 bp in Length | 868    | 0   |                    |                                   |
| 43802     | 44675: contig of 874 bp in Length | 874    | 0   |                    |                                   |
| 44676     | 45538: contig of 863 bp in Length | 863    | 0   |                    |                                   |
| 45539     | 46405: contig of unknown length   | 863    | 0   |                    |                                   |
| 46406     | 47269: contig of 864 bp in Length | 864    | 0   |                    |                                   |
| 47270     | 48139: contig of 870 bp in Length | 870    | 0   |                    |                                   |
| 48140     | 48999: contig of 860 bp in Length | 860    | 0   |                    |                                   |
| 49000     | 49866: contig of 867 bp in Length | 867    | 0   |                    |                                   |
| 49867     | 50729: contig of 863 bp in Length | 863    | 0   |                    |                                   |
| 50730     | 51588: contig of 859 bp in Length | 859    | 0   |                    |                                   |
| 51589     | 52460: contig of 872 bp in Length | 872    | 0   |                    |                                   |
| 52461     | 53323: contig of 865 bp in Length | 865    | 0   |                    |                                   |
| 53324     | 54189: contig of 866 bp in Length | 866    | 0   |                    |                                   |
| 54190     | 55053: contig of 864 bp in Length | 864    | 0   |                    |                                   |
| 55054     | 55910: gap of unknown length      | 864    | 0   |                    |                                   |
| 55911     | 56776: contig of 866 bp in Length | 866    | 0   |                    |                                   |
| 56777     | 57669: contig of 893 bp in Length | 893    | 0   |                    |                                   |
| 57670     | 58529: contig of 860 bp in Length | 860    | 0   |                    |                                   |
| 58530     | 59404: gap of unknown length      | 860    | 0   |                    |                                   |
| 59405     | 60263: contig of 859 bp in Length | 859    | 0   |                    |                                   |
| 60264     | 61140: contig of 877 bp in Length | 877    | 0   |                    |                                   |
| 61141     | 61997: contig of 857 bp in Length | 857    | 0   |                    |                                   |
| 61998     | 62885: contig of 881 bp in Length | 881    | 0   |                    |                                   |
| 62886     | 63756: contig of 871 bp in Length | 871    | 0   |                    |                                   |

|       |        |                            |
|-------|--------|----------------------------|
| 888   | 1780:  | contlg of 893 bp in Length |
| 1781  | 2645:  | contlg of 865 bp in Length |
| 2646  | 3525:  | contlg of 860 bp in Length |
| 3526  | 4401:  | contlg of 876 bp in Length |
| 4402  | 5290:  | contlg of 889 bp in Length |
| 5291  | 6155:  | contlg of 865 bp in Length |
| 6156  | 7151:  | contlg of 956 bp in Length |
| 7152  | 8036:  | contlg of unknown Length   |
| 8037  | 8948:  | contlg of 912 bp in Length |
| 8949  | 9836:  | contlg of 888 bp in Length |
| 9837  | 10705: | contlg of 869 bp in Length |
| 10706 | 11579: | contlg of 874 bp in Length |
| 11580 | 12459: | contlg of unknown Length   |
| 12460 | 13321: | contlg of unknown Length   |
| 13322 | 14190: | contlg of 869 bp in Length |
| 14191 | 15058: | contlg of 868 bp in Length |
| 15059 | 15923: | contlg of unknown Length   |
| 15924 | 16802: | contlg of unknown Length   |
| 16803 | 17660: | contlg of 878 bp in Length |
| 17681 | 18548: | contlg of 868 bp in Length |
| 18549 | 19433: | contlg of 865 bp in Length |
| 19414 | 20289: | contlg of 876 bp in Length |
| 20290 | 21156: | contlg of 867 bp in Length |
| 21157 | 22031: | contlg of unknown Length   |
| 22032 | 22916: | contlg of 875 bp in Length |
| 22917 | 23786: | contlg of 885 bp in Length |
| 23787 | 24655: | contlg of unknown Length   |
| 24656 | 25519: | contlg of 863 bp in Length |
| 25520 | 26383: | contlg of 864 bp in Length |
| 26384 | 27253: | contlg of 870 bp in Length |
| 27254 | 28119: | contlg of unknown Length   |
| 28120 | 28987: | contlg of unknown Length   |
| 28988 | 29861: | contlg of 874 bp in Length |
| 29862 | 30737: | contlg of 876 bp in Length |
| 30738 | 31601: | contlg of 864 bp in Length |
| 31602 | 32481: | contlg of 880 bp in Length |
| 32482 | 33347: | contlg of 866 bp in Length |

[illegible]

```

30139 CAGCATACGAGATGGGAGAAATTTTGGCATCTCTCATCTGCAGCAAGG 30090
1109 ythrlleglthlthleuylsuglgluyluylsyrthllysglnarqcyatyrg 1126
|||||
|||||
30089 GCTAATATCTCAGATCTCATCTCATGACTT.....AAACCAATTACCAAG 30046
1126 luytsglncgluylthlythlthlncgluylnglmcgluyluamln 1142
|||||
|||||
30045 AAAAAACCAACCCCATCTCCAAAAAGTGGCAAGAGGTT..... 30005
1143 lvsasnserserfvalproleumlqgluylsuglvalylsglnalaph 1159
|||||
|||||
30004 AGACCCAGCCATTTTCCAAAAAACCCATTTTATTCACGCCAACCACTTTT 29955
1159 egdluylsuglvalgllylletllysalaserleuarglulysglucug 1176
|||||
|||||
29954 AAAAAAAGAGCTC...ATCATCTAGCTGCTCATAGCA.....ATG 29917
1176 luserglnsrlsyrthylsglulglvalyserylsglncislerc 1190
|||||
|||||
29916 CAATTCAAAACCAATGAGATCATCTCATCTGCAGCATTTAGAA 29873

seq_name: gb_pat: AX193084

seq_documentation_block:
LOCUS AX193084 651 bp DNA PAT 15-AUG-2001
ACCESSION AX193084
VERSION AX193084.1 GI:15211039
KEYWORDS
human.
ORGANISM Homo sapiens
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catartida; Hominoidea; Homo.
REFERENCE
1. Classes 1 to 4411. Sreelaxi H., Benson D.R., Meagher M.J.,
Stokl J.A., Kling G.B., Wang T. and Jiang Y.
Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
Patent: WO 0149716-A 651 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. .441
/organism="Homo sapiens"
/db_xref="taxon:9606"
ORIGIN
127 a /db_xref="taxon:9606" 102 t 1 others
BASE COUNT 127 a 97 c 114 g 102 t
ALIGNMENT_SCORES:
Quality: 677.00 Length: 137
Ratio: 5.015 Gaps: 1
Percent Similarity: 98.540 Percent Identity: 94.161
Alignment_block:
US-09-750-590-2 x AX193084 ...
Align seg 1/1 to: AX193084 From: 1 to: 441
83 ThrsrAspThrlaIdlYAgsanaIdaIduHslsleuIdaIdaIdytrcgl 99
|||||
|||||
32 ACCGATGACGCGAGGAGGAATTCCTTCACGCTGCTGCTACATGATG 81
99 yHlslAluylsyrthylsuglncgluyluyluamlnlthlyranycyarprothcdln 116
|||||
|||||
82 ACATCATCTGTGCTCCAAAACCTTCATCAAGTCAATTTGTCCCACTGAC 131
116 lsvAlasrleuIdlglYrGththlAlaIduHslsAsrAlAlaIdaIdaAsp 132
|||||
|||||
132 ATCCAGACCTCGACGGAGGAATTCGATCTTCAGATCCGCCAATSCAGAT 181
133 CysrProSerSerThglncleuylsyrthylsuglvalyserylalaph 149
|||||
|||||

```

```

182 TCTCTTACATACAGCTGCTTGTGACATGGGCGCTCTGCAATGC 231
149 alvyspvalaspaiatgthgthproleuvalaleuathgthlmetcysa 166
232 CAGGATGATGAGGGGGGACACACTGTTCTGGCTACTAGACTACTA 281
166 tggprrthtgcysgthleuileuileuileuileuileuileuileu 182
282 GGGCAACAAATGTGTAACTGCTGATGATGAGAGAGGAGATTTATTC 331
183 ATGASPLYSGLAASNAATGThAlaLeuMeLeuAlaGlyGlyGlyCyl 199
332 GAGACAGAACAAACAAACAGTCCCTCACTGATGCTGCGATATGCTG 381
199 alvyspvalvslgylleuileuileuileuileuileuileuileu 215
382 CAGAGATCCAGTATGACTTAAATTAATAATGGTGTCTATKATNACTG 431
216 leuaspala 218
432 CTGCAATGC 440

seq_name: gb_pf:AB028997

seq_documentation_block:
LOCUS AB028997 5360 bp mRNA prt 04-AUG-1999
DEFINITION Homo sapiens mRNA for KIA1074 protein, complete cds.
ACCESSION AB028997
VERSION AB028997.1 GI:5689484
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
1 (siles)
REFERENCE
AUTHORS Kikuno,R., Nagase,T., Ishikawa,K., Hiroseawa,M., Miyajima,N.,
Tanaka,A., Kohani,H., Nomura,N., and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
X. The complete sequences of 100 new cDNA clones from brain which
code for 180 proteins in the
JOURNAL DNA Res. 6 (3), 197-205 (1999)
MEDLINE 99397452
REFERENCE 2 (bases 1 to 5360)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) to the DDBJ/EMBL/Genbank databases. Osamu
Ohara, Kazuo DNA Research Institute, Laboratory of DNA Technology,
1-5-32, Kiba, Kiba-shi, Chiba 292-0812, Japan
E-mail:ohara@dna.riken.ac.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914
FEATURES
source
1..5360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="h06637"
/cds_1=1..5360-coding
/cds_2=1..5360-pb1uescriptII SK plus"
/cds_3=1..5360-brain
/gene="KIA1074"
151..5280
/gene="KIA1074"
151..5280
/gene="KIA1074"
/codon_start=1
/product="KIA1074 protein"
/protein_id="BA03026.1"
/translation="MKTKSKGSPFLGSPNARRSGAGGGGEGGATYSQGTGIVD
RDGLIKHKAASGAVARVOOILLRNGLNDNRKRPPLAIVGNTALTYAVLIDVR
KCOLAVCDENRPLAKAVOCCEKCATILLGAGDNIPLAIVGNTALTYAVLIDVR
VATKLLDYANTLAKNKDILPILLAVSGKQKQAVPEFLIKKANVANVADLSSHOLI
SERBERIPHSQNSVDESDSLRSIRSGKGVDSWPTSDDEVDNPKRVKVP
SLAKMTAQSQSRKRLKLEATYGTIVGTRNTLFEEDSDSDSDVVEVLEPTTSIKVCS

```

```

BASE COUNT 2093 a 861 c 1177 g 1229 t
ORIGIN
alignment_scores:
Quality: 654.00 Length: 1588
Ratio: 0.766
Percent Similarity: 53.778 Percent Identity: 22.103

alignment_block:
US-09-750-590-2 x AB028997
Align seg 1/1 to: AB028997 from: 1 to: 5360
24 ArgLeuMeValGAlaAlaGluGlyAlaGluYValSerSer11 40
295 AAGATCCAAACAGCTGCCACCGGGATMTGGCGAAATGTCAGCAAT 344
40 e.....LeuAlaLysLysValAspProGlyLysLeuValAlaGlu 55
345 CTTTCTTCTCAAGCAAGATGCTTGAAAC.....GATGAGCAACAAATGA 388
55 LAlaSerAlaPheMetValAlaLysSerGlyGlyValLeuLeu 71
389 ACAGACAGGCTCTGATTTGGCTGTGGCTGATGATGATGATGATGATG 438
72 AsnAlaLeuLeuLeuLysGlyValAspIlePhePheAspPheAlaG 88
439 ACTCTCTGTGTGACAGAAATGCGACCTCAATGTCGTCAACCAACAA 488
88 yHtAspAlaLeuLeuLeuAlaLysLysValGlyLysAlaLeuValGlu 105
489 CAGGACAGCTGTGATGAGAGCTGTGATGATGATGATGATGATGATG 538
105 LAlaLysLeuLeuLeuLysValAspProGlyLysValAlaGluGly 121
539 CTATCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 588
122 AtGThAlaLeuLeuLeuAlaLeuAlaLeuAlaLeuAlaLeuAla 138
589 AACCTCTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
138 AlaLeuLeuLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAla 155
639 AAGCTGCTGTTGTATGATGCAAAATATTCAGCAAAACAAACAGATGAC 688
155 tghThProLeuValLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeu 171
689 TCACACACTTTTACCTTGATGATGATGATGATGATGATGATGATGATG 738

```

```

172 LeuLeuIleAspArgGlyAlaAspIleAsnSerArgAspGlyGlnAsnAsp 188
173 TTTTAAATTAAGAAAAGCAAAATGTAAATGAGTAAAGTAAATGCTGCTCAAA 1673
739 TTTTAAATTAAGAAAAGCAAAATGTAAATGAGTAAAGTAAATGCTGCTCAAA 788
188 gthAlaIleuMetLeuGlyCysGlyTyrGlyCysLeuAspAlaValGly 205
789 CACTCACAACCTAAATTCA...CAATAT...AAAGCAAAA 823
205 aLeuIleIleLys...AsnGlyAlaAspValThrLeu 215
824 GGTATCTCAAAACATCTCTCCAAAATAGTAAATAGTAAAGTAAAGCTCT 873
216 LeuAspAlaLeu...GlyIleAspSerTyr 225
874 GAAGACTCTCTTAACAGAGCTTCTGCCAAACGGGGCTGTGATGATCATG 923
225 TTYAlaArgIleGlyAspAsnLeuAspIle... 235
924 GCTCTACTCAACCAAGCAAACTTTTCAATTTTCAATCAAGATGCCCA 973
236 ...LeuThrLeuLeuIleuThrIleSerGlyAsnSerAsnLysGly 249
974 AACCAAGCTTCAACCAAGCAAACTTTTCAATTTTCAATCAAGATGCCCA 1023
250 ArgGluLeuThrPlyLysGlyProSerLeuGlnIleAspAsnLeu... 264
1024 TTAAAGCAACATATGCACTGTGACAAAGCAAAATACAAATGCTGCTTGA 1073
265 ...SerGluMetLeuAspGlyAlaAspThrLysSerAsnGln 278
1074 GGATAGACATTCGATGATGCAAGATGAAATGCTGCTGTTGAAGCTTCTCT 1123
278 PGLuIleGlnAsnIleGlnAspLeuGlu...IleGluAsnGly 291
1124 CAACATCATCAAAAGCTAGCTGCTTTTCTCATCTCACTCACTCACTCACT 1173
292 AspLeuLysGluAsp...LeuAsp 298
1174 GACCTCTCTCAAAAACCTTCCCAACAGTCTTACCAAAACCTGCTCTTAT 1223
298 GlyAlaIleGlnIleGlnIleLeuLeuAspLysValAsnGly... 313
1224 GAGCAAGAACCAACCAAGCAAACTTCCCAACAGTCTTACCAAAACCTGCTCT 1273
314 ...LeuGlnIleGlnIleAsnGlnIleGluIleAsp 323
1274 ATATATGAAAGTCTCTCACTCACTCACTCACTCACTCACTCACTCACT 1323
324 ValAlaAspAspLeu...GluSerGlnLysGlnLysLeuLysSerLeu 339
1324 TATCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1373
339 ValAlaIleLysGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 356
1424 TCTGTAAATATTTCCCAAGCAAGTATGCTGCTTCTTACCTGAGGCTGCA 1473
373 HisPheArgLysGluAspLeuLeuLysGlnIleGlnIleGlnIleGlnIle 389
1474 GACGAAAGCAAAATATATGCAAGTATGCAAGCAAGCAAGATGCTTTA 1523
389 HisPheArgLysGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 395
1524 TATACCTCTCTTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 1573
396 ...ThrGlyMetProVal...HisMetGlnSerAsp 406
1574 AGCATACAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1623
407 Met...LeuAlaPro...LeuGluLeuAlaLeuProAsp 417
1624 AATATCTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1673
417 nGlnIleAspSerTyrSerGluAsnGlnIleLeuLysGlyGluLeuIleValIle 434
1674 TAAAGCAGGCAAGTAAAGCAATGATGATGATGATGATGATGATGATGATGAT 1723
434 eGArgThrPheCysAspSerAlaLysGlnAspArgLeuLysGlnIleGlnIle 450
1724 ACTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1773
451 GlnLeuAlaIleLysValAlaIleGlnIleLysGlnIleGlnIleGlnIle 467
1774 AACCAAGCAAG...GTTCAAAACCA 1796
467 ValArgLysGlnIleAspSerArgGluGln...IleLysGlnIleGlnIle 482
1797 AAGCAAAACCAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1846
482 sPAlaLeuLysAspValGlnIleLysMetTyrIleSerGlnIle... 496
1847 ATGCTGCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1896
497 ...LysValLysGlnIleMetGlnIleHisPheLeuAlaIleLysGlnIle 512
1897 CAABAAGCAAGCTGCAAGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1946
512 ThrSerAspAlaIleThrIleAsnIleAspIleMetGlnIleLysGlnIle 529
1947 AAATTAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1996
529 sPLeuLeuLysAspMetLysLysLysTyrGlnIleLysAspIleGlnIleVal 545
1997 GCACGAAAGCAAAACGACCTGCAAGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 2046
546 ...GlyLys...LeuAlaAsnGlnIleLysGln...AsnI 556
2047 TTTGGAAAGCAAGCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2096
556 uMetLeuValGlnIlePheArgAspIleGlnIleLys... 568
2097 CAGCAGTTTAAGTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2146
569 ...LeuMetGlnIleAsnLysArgLeuGlnIleLysGlnIleSerMet... 582
2147 CATCTAATGAAAGCAAGCTGCAAGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 2196
583 ...CysGlu 585
2197 GTTGAATGATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2246
585 uGluArgGlu...LysArgIleArgLysLeuThrIleuMetGlnIle 599
2247 ACCCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2296
600 ...GlnLeuLysAspSerAlaIleLysLeuAlaLeuSerIleProAsnIle 616
2297 TGGATGTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2346
616 LysPheGlnAsnMetLysSerLeuLeuSerAsnIle...LeuAs 629
2347 TCATGTAAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2396
629 nGluLysAlaLysLysLeuIleAspValGlnArgGlnIleGlnIleGlnIleGlnIle 646
2397 AGTAAGAAATTAAGAAATGCAAGCAAGCTTAAATGATCACTCACTCACTCACTCACTCACT 2446
646 eAsnGlnIlePheArgProLeuLysArgGlnIleGlnIleLysLeuLysLysLys 662
2447 TATCTGAACCAAGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2496
663 LeuAlaGlnIleVal...LysProGlnIle 671

```



```

2497 TGGAGACGAGACGTGCTCTTTGAGATTACGTTAAACAGAGAGAA 2546
671 u.....HISGLINLEULYSERARGLEUC 680
2547 GAAGAGAAAGAAATGCTGATACGTTGATGAAAAATAGGACAGTTAA 2596
680 IUGLILYSERGLYLEUENGLYLSARGLE.....THIRGLEU 693
2597 GAAGAAAGAGAGCAATATAGCAAAAGCTGAACTGAAACAGACGTT 2646
694 THIRGLYSASNGINLHLEUGLINSGLIIEGLILYLSVALYLSLEUAS 710
710 PASHLYLEULeuThrGLINGINVALASNALEULeuThrGLIUEULYAS 727
2689 .....AAAAGTAATTTGAAATACGCTGTTCAAGCCAGAAATG 2725
727 snVALPROLEULYSVALSERGLUJLUEULYSERHISASPAVALIE 743
2726 ACCTCGACAGCGCACTTCTCGAGACAGCAATGCC..... 2760
744 VALASPARLEULenLYLEULYSERASP.....VALTHIRHISLY 757
2761 .....AGATGTTACAGATGATGSAATTCGACACATGACCT 2795
757 sTYrThrGLILYLSLEUGLJUEC.....GLILYLSLEULeuGLIUA 772
2796 /TTCACAAACAAAGACATTCGAAATGGCTGAAAGAAATGATTCGTAGA 2845
772 snALE.....SERLEUSERTYASNAVALSERARGLEUGLUTHVAL 785
2846 ATTCTCATGTCATGAGAGAAAAAGACCTTCGCAATAAATAGCATG 2895
786 PHEILEPROGLIUGLIRHISGLILYSGLJUEULeuGLIALeULYSERAS 802
2896 TGG.....CAGGAGCAATTTCTCATGCTTAAGCATAGA 2927
802 nILEPHEGLILEULYLSGLILEUSERTGLILEUASNALEULYSGLYC 819
2928 AATAACACATATAAAATCAACCGAGAAAGAAAGAAAGAAATGTTTG 2977
819 IUAASPGINGLILYLSILEYrSERLEUWELSERGLIUSASNAAPLEULYS 835
2978 AGGAC.....CTTAAATTTGMAAGAAAGAAAGAAAGAAACCTTCAG 3018
836 IYSHIRMETSERHISGLINTYVALPROVALYLSERHISGLIUGLILY 852
3019 AAGACT.....ATADACACAGATGACAA..... 3042
852 sPHIRALEUSERTYRLEUASPELYLSHIRASNAAGLILUVALASPY 869
3043 .....ACATTAACACAAACAAATTTCCAGATTAATAGCCCTTAGGTTC 3088
869 ALLYLSYLSYCSGLIUSPILeASNGINGLUPHEVALYLSILYASAP 885
3089 TGCACGCTGACATGCAATGCTAAATTCACACGAGAAAGAAAGAAACAA 3138
886 GLIUAINGLILILEULYASNALEUGLINSNAHTRHISASNGINVALY 902
3139 ACCAAGGAAAGACTGGAAGACAGATTGAATCACTTCACATGATGCC 3188
902 snALA.....GLUTYLILESERLEUARGLUNISG 912
3189 TGTCTGTATACATGATCGTATCAAAATGACAAATGACAAAGAAACATAG 3238
912 IUGLILYWEISERGLYLEUARGLYSERMELTYLSVALIDIASPARS 928
3239 AACCTTCGTTCCAGAGCAAGAGATGAATTCCTCGCTTAACAGCAAA 3288
929 SERIALGILILEUENGLYSTTYLSYLSERGLINGLILUVALTH 945
.....
3289 ATGAATTTGATGTGTCTAACCTAAAGATACATAGAG.....AT 3329
945 ILEUHLIEGLIUGLILIEALALAGLILYASARGLILEUASPHIRHIEGLIN 961
3330 TCTTCTTCACACACTTTTAAACCTGAAAGAAATCTCATACAGCTGAA 3379
962 .....GLUCYALILEYLSLEULYTYRVALPROLIE 971
3380 TTGAATTCATACACAGAGAGATGCCCTACAGAAAGAG.....ACT 3420
972 ILESERTLUGLILUCYSGIUGLIRGLYSPHELYALATHIRGLIUSGLULE 988
3421 TTGGGTTTGAACGGTGTACAAAGACCTACAGCAACACAGCTGCAT 3470
988 ULYSGIUGLILeuSERGLINLHTRHISGLINLYSTYRASNTHIRSERGLIUG 1005
3471 GAAGGAA.....ATGGAAGAAACAAATGTATCAAAATGACACAGTTA 3508
1005 IUAIALYLSYCS.....LYSGINGLILUASN.....ASP 1014
3509 AAGCTGATTAATTCATTCGAAAGAAAGAAAGAAATGCTTGACAGCAATGCTCT 3558
1015 LYSLEULYLSGLIULeULeuThrGLIUEUGLILYASNAAPLEULYAS 1031
3559 CAACTCAAAAGTGAAGATATGCTCTTGCAACAAACAGTGCATGCCA 3608
1031 nVALHISLIEGLIUSASERTYrGLIUTHGLI..... 1041
3609 CAAACAGCTGACATTAAGAAAGAAAGAAAGCAATGATTAATTCACAGCCAGT 3658
1042 .....ARGALILEUSERTARGLYS..... 1047
3659 TTCATGCTATTGTGCAGAAAACCTTAAGCTGAGAGTGAAGAAAGCAATCTT 3708
1048 ...ThrGLIUGLILUASNARGLIN.....LEULY 1056
3709 CTGCTGAAAGAAAGAAATAGAGAGCTTAATTCAGATGATTAATCACTTAA 3758
1056 SASPLEULeuGLINLYSTYrThrGLIUALIYSGIULYLSGLIULYLSLEU 1073
3759 AGAAAGACAGATACATATGAAATTAAGAAAGCAAGCAAGAAAGATGTTG 3808
1073 ALIGLILUASNALYSGLINHrSERGLIULEUVALALGINTHIRLEU 1089
3809 TGAACACACTTCACACAGAACTACCTGATCCCTTAAAAAACAATCTATG 3858
1090 LEULIN.....LYSGINLISVALPROLEULU..... 1098
3859 TGCAGAGCTTCACCTGAGCTTACGCTGATCGATTAATTAATTAAGGA 3908
1099 GLIUALIGLISERLEULYLSYSERLEUSERTGLYTHIRHIEGLIUTHLEU 1115
3909 TGAACACAGAGATTAAGAAAGAAATTA.....GGTCAAAATCGAAATCAAT 3955
1115 YSGIUGLILeuLYSHIRGLIUSGLIN.....ARGCTYrGLIULYSG 1127
3956 TGCACAAACACAGAGATCGACATTCACAGAACTGCATGATGCTGTAGAA 4005
1128 GLIUGINGLIRHVALTHIRGLILEUASNGLJUEULeuGLIUAENGLIULYAS 1144
1144 nSERSERVALPROLEULINGLILUHLISLEUGLILYSGIUALPHEGLIU 1161
4041 TCCCAAGTTAAAGTTACAGTCAAAAGCAAAATGSAACAAATTTGAGAGAG 4090
1161 YSGIUALYALGYIIELELYALSERLEUARGLILYSGIULINGLILUSER 1177
4091 TTCAAGAAACCTGTAAATCAATATTTGCTGAAAGATGAAAGAAAGCAAA 4160
1178 GLIASNLYSHIRGLIULYASER..... 1185
4141 TTAAGAAACTATATGCAATTAAGACAGTCACTGCAATGTAAATTTGATCA 4190

```

```

1186 .....LysLeuGlnSerLeuIleGlnThrLysG 1196
1191 AGAATGAGAAAAATGTTGAAATTAAGAGAGAGAGATGCTGATTAAG 1208
1196 LAlaLeu.....LysLeuGlnThrArgGluValValAsp 1208
1201 .....LysLeuGlnThrArgGluValValAsp 1208
4241 ACCCTTAAAAATGACAGAAAGACGTTAAATGATATGAAATGAGAA 4290
1209 LeuSerLysArg.....LysAlaThrLysSerAspLeuGlnThrG 1222
1211 .....LysAlaThrLysSerAspLeuGlnThrG 1222
4291 TTTAGTTCTCATGAGCATTTAAAGACTGACATTTGAAATGATGATTC 4340
1222 nIleSerAspLeuAsnGluLysLeuAlaAsnLeuAsnArgLysArgLys 1239
1227 .....LysAlaThrLysSerAspLeuGlnThrG 1239
4341 GATTAAATACCTAAAGCATTAAGATGATGATGATGATGATGATGATG 4390
1239 LuValCysGlnGluValLeuAlaLysLysLysGluLeuSerAlaLys 1255
1241 .....LysAlaThrLysSerAspLeuGlnThrG 1255
4391 CTCACAGTCTCAAGTCTCAATCTGATCAAGAGAGAGAGAGAGAGAG 4437
1255 ArgCysLysGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1272
4438 ...CAAG 4466
1272 nGlnGluArgCysAspLysSerLeuThrThrLeuGlnLeuGlnArg 1289
1277 .....LysAlaThrLysSerAspLeuGlnThrG 1289
4467 ACNAAAGAAATGTCNAAA..... 4485
1289 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1305
4486 ..CTACAG 4524
1306 GluLeuLeuAsnAspValGluArgLeuLysGlnAlaLeuAsnGlyLeu 1322
4525 AACCTGAGAGTCATATAGAAAGAGATGATGATGATGATGATGATGAT 4574
1322 nGlnLeuThrArgLysSerProSerLysArgGlnSerGlnLeuI 1339
4575 ACAG.....TATAACAGAGATGAGAAAGAGAGAGAGAGAGAGAG 4618
1339 LeuSerLeuGlnGlnValArgSerLeuGlnGlnGlnAlaAsp 1355
4619 CAGAAAAATTAAGAGTCATCTATTTACAGGACAGAGAGAGAGAGAG 4668
1356 AlaAspArgGlnHisGlnGlnValIleAlaIleArgThrHisLeuLe 1372
4669 .....CAGAAACCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4703
1372 uSerAlaLeuGlnGlyHisMetAsp.....GluAspValGlnAla 1386
4704 TCGTTCATGAAAGAGTCATGATCACTCAAGATTAAGAGATCTGAACTG 4753
1386 LeuLeuGlnIle 1390
4754 AACCTCCAAATA 4767

```



Align seq 1/1 to: MAV8134 from: 1 to: 2322

```

675 LeuLysSerArgLeuGluGlnLysSerGlyGluLeuGlyValArg1LeuH 691
      ::
2 GTTAAGACGACATTAGACAGAAATACAGAGAACTGGCAAGAAATACAC 51
691 rGluLeuThrSerLysAsnGlnThrLeuGlnLysGluLeuLysValC 708
      ::
52 TGGGTAAACATTGAAATACAGACACTACAGAAAGGAAATGAAAGTTT 101
708 yAlaAspAsnLysLeuLeuThrGlnGlnValAsnAsnLeuThrThyGlu 724
      ::
102 ATTGTGATTAATACGCTCTCCACAGCAGCACTCAATTAACAATTGAA 151
725 MetLysAsn.....ValProLeuLysValSerGluLysLeuLysSer 735
      ::
152 ATGAAAGAAATCAATATTCTCTCTTAAGATACAGTAAAGACAAAGATC 201
735 rHisAspValLysValAspAspLeuLysLysLysSerArgValThrH 756
      ::
202 ACATGATGCATATTGATGATCTTAAATAGAAAGCTTTTGGATGAACAC 251
756 lAlaSerLysSerLysAsnValSerArgLeuGlnThrValHisThrProP 789
      ::
773 AlAspSerLysSerLysAsnValSerArgLeuGlnThrValHisThrProP 789
      ::
302 GACAGCTTAAGTAAGATCAACCGCCTAGAAACGTGTTGTGACTCC 351
789 oGluLysArgLysGluLysGluLeuMetAlaLeuLysSerAsnLysThrGlu 806
      ::
352 TGGGAACATGAAAGAGATAAATAGCTGAAATCAATATTGTTGAC 401
806 gLysLysGlnLysSerGluLeuAsnLysLysGlyGluLysAspGlnGlu 822
      ::
402 TTAAAGAACGCTGTCTACGATTAAGAAATTAATGCTGAAAGCAAGAG 451
822 yLysLysLysSerLysSerGluLysAsnAspLeuLysLysThrMetSer 839
      ::
452 AAAATACACGCTCTCACATCGAAACACCTGAACCTTGAAGAACTGATGAG 501
839 rHisGlnThrValProValLysThrHisGlnGluLysLysThrAlaLeu 856
      ::
502 TTAATCAGATATGCTCAATTAAGAACCATTAAGAGAGTAAACACACCA 551
856 eArgSerThrLeuAspLysThrAsnArgLysLysLeuValAspValLysLys 872
      ::
552 ATGACAGCTTACCCAAACACTACAGAGAAATTTAAGATGTAAGAAAAA 601
872 CysGluAspLysLeuAsnGlnLysLeuLysLysLysAspLysAsnGln 889
      ::
602 TTTCGAAACATTAATACAGAAATTTGTAAGAAATTAAGATTAAGATTAAT 896
889 eLysLysValArgLysLeuLysLysAsnThrGlnAsnGlnValLysValLys 906
      ::
702 TCAGCTCTGGACAGACACAGAGCAAGAGTGGCTGCAATGCTACGAGAC 751
906 lAspSerLeuAspGlnLysGlnLysLysSerGlyLysLeuArgLysSerMet 922
      ::
752 AGAAAGCTGCAGATAGTAATGCTGAAATCTTGCCCACTACAGAAAGG 801
922 LysLysValGlnAspAspSerAlaGlnLysLysLeuValLysThrLysLys 939
      ::
939 rGlnGlnLysLysLysValThrLeuHisGlnGluGlnValAlaLeuAlaLysArg 956
      ::
802 CCAGAGAGAGATTTGTGACACTGACCGCAAAATTAAGCCAGAGAAAGG 851
956 lLeuLeuAspThrLysGlnGlnLysLysLysLysLysLysLysLysLys 972
      ::

```

```

852 AGCTGCACACATACAGAAATATCATTAAGTAAATATATGCCCAATTGTC 901
973 SerLeuGlnGluGlyGluValArgLysPheLysAlaThrGlnLysGluLeuLys 989
      ::
902 ACCTTTGAGAGATTCGAGAGAGAAATTTAAACAGACAGACAGACATAA 951
989 sGluGlnLysSerGlnGlnThrGlnLysSerAsnThrSerGlnGlnGlu 1006
      ::
1006 lAlaLysLysCysLeuGlnLysAsnAspLysLeuLysLysGluLysLeuThr 1022
      ::
1002 TCAGAAAAAACACAGAGAGAAATACAGAGTAAATTAACAGAGATTTTAC 1051
952 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 992
      ::
1023 LeuGlnLysAspLeuLysAspLysAsnValHisLysLysSerThrGln 1039
      ::
1052 CTTCAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101
1039 uThrGlnArgLysLeuSerArgLysThrGlnLysLysAsnArgLysLeu 1056
      ::
1102 AATGAAAGACATTAATACAGAAAGAAACAGACAGCTTAACAAACATGTA 1151
1056 yAspLeuLeuGlnLysArgThrGlnLysLysGlyGluLysGluLeu 1072
      ::
1152 AGACCTTGTCACAGAAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
1073 ValGlnGlnLysAspLysLysGlnLysSerGlnLysLysValAlaLys 1089
      ::
1202 GTAGAGAAATATGCGAAACAGACTTCTGAGATCTTGCAAGTCAAGTCT 1251
1089 vLeuGlnLysGlnHisValProLeuGlnGlnValGlnLysLeuLysLys 1106
      ::
1252 TTTCGAAAAACAGACATGTCATTTGAGACAGTTGAGAGCTTGCAAAAT 1301
1106 eLysLeuSerLysThrLysLysLysLysGlnLysLysLysLysLysLys 1122
      ::
1302 CTCTTAATGCGAATTTGAAATCTTAAGAGAAACACTGAGATGATGCA 1351
1122 rLeuGlnLysGlnHisValProLeuGlnGlnValGlnLysLeuLysLys 1139
      ::
1133 ArgCysArgLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1156
1352 AGGTGTACGAGAAAGACAGACAGACAGACAGACAGACAGACAGACAG 1401
      ::
1401 uGlnLysGlnLysAsnSerSerValProLeuGlnGlnValGlnLysLeu 1156
1139 GAG 1451
      ::
1156 yGlnLysAlaPheGlnLysGlnLysValGlnLysLysLysLysLysLys 1172
1452 AGAAGCATTTGAGAAAGAGTGGATTAATTAAGAAACCCACTGTGAGCA 1501
      ::
1173 LysGlnGlnLysLeuGlnLysAsnLysThrGlnLysLysLysLysLys 1189
      ::
1502 AG 1551
1189 rGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1206
      ::
1552 GAGAGGTGAGATTCGAAACAGACATTAAGAAATTAAGATGAGAGAG 1601
1206 AlValAspLeuSerLysArgLysAlaThrLysSerAspLeuGlnThrGln 1222
      ::
1602 TATGATCTGCTGTAATTAATTAAGAGCAAGAGAGAGAGAGAGAGAGAG 1651
1222 lLeuSerAspLeuAsnLysLysLysLysLysLysLysLysLysLysLys 1239
      ::
1652 ATTTCTAGCTTAATTAAGAAATTTGGCCCAATCTGATTAAGAAATGATG 1701
1239 vValCysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLys 1256
      ::
1702 AGTATGTAGAGAGAGTGTGTGATGCGAAAGAGAGAGAGAGAGAGAGAG 1751
1256 sPheLysGlnLysLeuLysLysLysLysLysLysLysLysLysLysLys 1272
      ::
1752 ATGAGAAAGAGATTAATGATCTTACGATTTACGATTTGAGAGAGAGAG 1801

```



```

651 CTGGCAATGTCAGACATGAGAAAGCTGCGAGTATGATCTGGAAATCTG 700
933 GAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 949
701 TGGCCACTGACGAAAGGCCAAGAGGATTTGGACCTGCTCCGAA 750
950 TTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 966
751 ATTTAAAGCCCAAGAGGACCTGACCAATACCAAGATCTTAAAGGT 800
966 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 983
801 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 850
983 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 999
851 CACGAGAAAGACACTTAAAGACGATTTACGAGCGACCAAAAGAT 900
1000 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1016
901 AGT 950
1033 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1049
1001 TCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1050
1050 GATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1066
1051 GAGCTAAACAAACAGCTTAAAGACTTGTACAGAAATACAGCAAGTAA 1100
1066 ATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1083
1101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1150
1083 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1099
1151 TACCTGACGTCGAAATCTTTGCAAAACACATGCTTCATTCGAAACG 1200
1100 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1116
1201 GTTGAAGCTGTGATGATGATGATGATGATGATGATGATGATGATG 1250
1133 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1149
1251 ACAAACGAAAGATGACCAAGGCTTTACGAGAAACACCGCAAGTGA 1300
1133 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1149
1301 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1350
1150 ATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1166
1351 GCAGACGATTTGCAATTTAAACAGCATTTTGAACAAAGTTGGAATCAT 1400
1166 ATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1183
1401 AATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1450
1183 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1199
1451 AAGCTTCGAAACTTCGTCGAGAGCTTCAAACTAAACCAACCTTAA 1500
1200 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1216
1501 AATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1550
1216 ATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1233

```

```

1551 AAGTATTTGGAGACAGATTTAGCTTAATGAAATTTGGCCATC 1600
1233 GATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1249
1601 TGAATGAAAGATGATGATGATGATGATGATGATGATGATGATG 1650
1250 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1266
1651 AAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1700
1266 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1283
1701 GCAAGAAATTAAGATGATGATGATGATGATGATGATGATGATG 1750
1283 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1299
1751 TCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800
1300 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1316
1801 AAGTATTAAGATGATGATGATGATGATGATGATGATGATGATG 1850
1316 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1333
1851 GATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1300
1333 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1349
1901 AAGGACGAGCCAGCTGATGATGATGATGATGATGATGATGATG 1350
1350 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1366
1951 GATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2000
1366 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1383
2001 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2050
1383 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1399
2051 TATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100

seq_name: /SIDS2/gcgdata/genomeseq/geneseq/ML1999.DAT.AAX40197
seq_documentation_block:
ID AAX40197 standard: DNA: 1980 BP.
XX
AC AAX40197:
XX
DT 02-JUL-1999 (first entry)
XX
DX
XX Sequence of C3Y5 gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
prostate cancer; ss.
OS Homo sapiens.
XX
PA W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998. 98W0-US14679.
XX
PR 22-JUN-1998. 98US-0102322.
PR 10-OCT-1997. 97US-0861564.
PR 10-OCT-1997. 97US-0061599.
PR 10-OCT-1997. 97US-0061765.
PR 10-OCT-1997. 97US-0948705.
PR 11-OCT-1997. 97GS-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.

```







```

101 CTCAGACGCTCAAAACGAGAACATGAAACAGCTTAAAGCACTTAAAC 150
681 GlnlySergLysGluLeuGlyLysArg1LehrGluLeuThrSerLysAs 697
151 CAGAAATTTAAAGCAAGAACTGGGAGAAAGATCTGATTAACATTAAGAAA 200
697 nGlnThrLeuGlnLysGluLysGluLysValLysLeuAspLeuLeu 714
201 TCAGACACTCAAAAGGAAATTTGAAATTTATTTGATTAATAGCTCC 250
714 eUrThGlnGlnLysAsnLeuThrThrLysLysAsn.....Val 728
251 TCAGAGAGACAGCAACCAATTAACATTAAGAAATTAAGAAATTAATGTT 300
729 ProLeuLysValSergLysGluLeuLysLysSerHisAspValLysVal 745
301 CTTTTAAAGTAACTGAGACATGAAATTAATGATTAATTAATTAATTA 350
745 PaPLeuAsnLysLysLeuSerAspValThrHisLysThrGluLysL 762
351 TGATCTTAATAGAAAGCTTTAGATGTAACAAAAATATTCGAAAGAA 400
762 YsLeuGlnLysGluLysLeuLeuGlnLysAsnLysSerLeuSerLysAsn 778
401 AGTTGAAATGGAAGAAATTTCTACGAAAGAAAGCACTTAAGTAAGAT 450
779 ValSerArgLeuGlnThrValAlaPheLleProProGluArgHisGlu 795
451 GTAGACCCGCTAGAAAGCTGTGTTGTACTCTCTGAGAAACATGAAAGAA 500
795 UkeLysLysLysLysSerAsnLLehrGluLeuLysGluLeuLeuSerg 812
501 GATATATGCTGAAATTCAGATTCGTAAGATGTAAGCAAGCACTGCTG 550
812 LysLeuAsnLysLysCysGluAspGlnGluLysLysLysSerLeuMet 828
551 AACTTAAGAAAAATGTGTAAGACAGCAAGAAATTCACGCTCTGACA 600
829 SerGlnAsnAsnAspLeuLysLysThrMetSerHisGlnLysValPro 845
601 TCGAAAGCACTGAACTGAAAGATGATGATGATGATGATGATGATGAT 650
845 LysThrHisGlnGluLysLysThrAlaLeuSerSerHisLysPhe 862
651 TAAACCCATGAGAGAGCTTAAATGACATGAAATGACATGAAATGAC 700
862 hTrnAsnArgLysValAspValLysLysLysCysGluAspLysGln 878
701 CTGAGAGAGATATTTAAAGATGAAAGAAATTTGTAAGAAATTAATAC 750
879 GluPheValLysLysLysAspGlnLysGlnLysLysValArgAsn 895
751 GAATTTCTAAAAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 800
895 uAsnThrGlnAsnGlnValLysAlaGluLysLysSerLeuAspGlnLys 912
801 AAGACCTGAGAGACCAATTAAGAGCTGATGATGATGATGATGATGAT 850
912 LysLysMetSerGlnLysLysArgLysMetLysLysValLysAsn 928
851 AGCCAAAGATGACCTGCTAAGTCAGAGCATGAGAAAGAGCTGACAGAT 900
929 SerAlaGluLysLeuLysLysLysLysSerGlnGlnLysLysLys 945
901 AATGCTGAAGATCTTGCCCAATGACAGCAAGAAAGCAAGAAATTAAT 950
945 rLeuHisLysGlnGluLysAlaGlnLysArgLysLysLysLysLys 962
951 ACTGCAATGCCCAATTAAGCCAGAAAGAGAGCTGCAATCAATTAAG 1000
962 LysGlnLysLeuLysLysLysLysLysLysLysLysLysLysLys 978

```

```

1001 AATGCAATTAAAGCTAAATATATGCCCAATTTGACGCTTGAAGAGTCGAG 1050
979 ArgLysPheLysLysLysLysGluLysGluLysGluLysLeuSergLys 995
1051 AGAAATTTAAAGCAAGAGAAAGAACTTAAGAAATTAAGATTAAGAGCA 1100
995 nThrGlnLysLysAsnThrSerGlnGlnLysLysLysLysLysLys 1012
1011 GACACAAAGATATGATGCTGACAGCAAGAAAGCAAGCAAGAAAGCAAG 1150
1012 LysAsnAspLysLysGlnLysLysLysLysLysLysLysLysLys 1028
1151 AGAATGCAAGCTTAAGAGAGAGAGATTTTACCTTCAGCAAGATTTGAGA 1200
1029 AspLysAsnValHisLysGlnLysSerThrLysThrGlnLysArgLys 1045
1201 GATTAAGCAGTTCTGATTTGAGAGCTCTCAATTAAGTAAGTAAGCA 1250
1045 rArgLysThrGlnGluLysAsnArgGlnLysLysAspLeuGlnLys 1062
1251 CAGAAAGACAGCAAGCACTTAAGCAAGCAAGCAAGCAAGCAAGCAAG 1300
1062 YrThrGlnLysLysGlnLysLysLysLysLysLysLysLysLys 1078
1301 ACAGCAAGTAAAGAAATGTTAAAGAAAGCAAGCAAGCAAGCAAGCAAG 1350
1079 GlnThrSergLysLysLysLysLysLysLysLysLysLysLysLys 1095
1351 CAGACCTTCTGAGATCTTGAGAGCTGCAATTTCTGCAAAAGCAAGCAT 1400
1095 LProLeuGlnGlnValGlnSerLeuLysLysLysLysLysLysLys 1112
1401 TCCATTTGCAAGAGCTTGAAGCTGCAAAATTTCTTAATGSCCAATTC 1450
1112 LysThrLeuLysGlnLysLysLysLysLysGlnArgCysTrpGlnLys 1128
1451 AATATCTTAAGAGAAAGCACTTAAGCAAGCAAGCAAGCAAGCAAGCAAG 1500
1129 GlnGlnThrValThrGlnLeuArgGlnLysLeuGlnLysGlnLys 1145
1501 CAGCAGACAGTGAACCAATGCAATTTGTTGGAATGCAAGCAAGCAAG 1550
1145 rSerValProLeuAlaGlnLysLeuGlnLysLysLysLysLysLys 1162
1551 TTCTGTACCCCTGACAGCAATTTGATTAAGAAAGCAATTTGAGAGAG 1600
1162 LysValLysLysLysLysLysLysLysLysLysLysLysLys 1170
1601 AAGTTGCAATCAATTAAGCAAGCAAGCTG 1626

seq.name: /STD2/9c9data/geneseq/geneseq/NA2001.DAT:AAH03594
seq.documentat_block:
XX AAH03594 standard; cdna: 848 BP.
XX AAH03594;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cdna clone (5'-primer) SEQ ID NO:429.
XX
XX Human: primer: detection: diagnosis: antisense therapy: gene therapy: ss.
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000: 2000EP-0116126.
XX
XX 29-JUL-1999: 99JP-0248036
XX 27-AUG-1999: 99JP-030053.

```



XX tumor; infection; depression; proctitis; ss.  
 XX Homo sapiens.  
 OS NC0200021991-A1.  
 PN 20-APR-2000.  
 PD 15-OCT-1999; 99MO-US24206.  
 XX 15-OCT-1998; 98US-0104436.  
 PR (GENY) GENETICS INST INC.  
 PA  
 XX Jacoba K. McCoy JM, Lavalie ER, Collins-Ragle LA, Evans C;  
 PI McReery D, Treacy M, Bowman MR;  
 XX MPI: 2000-317938/27.  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (ESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 PS Claim 1: Page 616: 803pp; English.  
 XX  
 CC AAA3426 to AAA45925 represent specifically claimed secreted expressed  
 CC sequence tags (ESTs), isolated from human, mouse, chicken and rat  
 CC tissue sources. The ESTs are useful as probes for the identification of  
 CC the tissues they were isolated from. The activities include:  
 CC chemokine; proliferative; immunomodulatory; haemostatic;  
 CC cytoskeletal; antibacterial; antifungal; antiviral; antiinflammatory;  
 CC antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;  
 CC neurotropic; antiparkinsonian; antiproliferative; cerebrotective;  
 CC photoreversant; and antidepressant. The ESTs can be used for gene  
 CC identification and isolation of full length cDNA.  
 CC molecules which correspond to the ESTs. Proteins encoded by the ESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
 CC disorders (hemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours bacterial (HIV), anti-infectious agents and  
 CC proestasis. AAA45926 to AAA45931 represent linker variants which are given  
 CC in the exemplification of the present invention.  
 XX  
 XX Sequence 571 BP; 246 A; 80 C; 125 G; 120 T; 0 other;  
 alignment-scores:  
 XX Quality: 858.00 Length: 183  
 XX Ratio: 4.714 Gaps: 0  
 Percent similarity: 99.454 Percent identity: 94.536  
 alignment\_block:  
 US-09-750-590-2 x AAA45014  
 Align seg 1/1 to: AAA45014 from: 1 to: 571  
 1133 ThrGlnLeuArgGlnMetGlnGluAngGlnGlyAsnSerSerValProLeu 1149  
 23 ACCAACTGTCATCAATGTGGAGATCAAAAGACCTCTCTGATCCCT 72  
 1149 uAlaGluHisLeuGlnGluValylGluAlaPheGluGlyGluValGlyLeu 1166  
 73 GCGAGAGCAATTCGAGATTAAGAGAGCATTTGGAGAAAGAGTTGGAAATCA 122  
 1166 IeIyGAlaSerLeuArgGluGlyGluGluGluGluGluGluGluGluGlu 1182  
 123 TAAAGCGACGCTTGAGAGAAAGAGAAAGAAAGCCAAAACAAATGGA 172

1183 GluValSerIleuGlnInserGluIleGlnAsnThrIysGlnAlaLeu 1199  
 173 GAAGTCTCCAACTTCACCTGCGGAGGTTCAGAAATCAACAGCTTTAA 222  
 1199 sIySLeuGluThrArgGluValAlaPheLeuSerIySylsAlaThrL 1216  
 223 AAAATTCAGACATGAGAGAGCTTGCATCTCTTAAATAAAGCAACA 272  
 1216 ySerAspLeuGluThrGlnIleSerAspLeuAsnGluIySLeuAlaAsn 1232  
 273 AAGGCAATTTGGAGACACACATCTTCCTTAAATGTAAGAAATGGCCAG 322  
 1233 LeuAsnArgIySylsGluGluValCysGlnGluValLeuAlaAlaIySyl 1249  
 333 CTGAATGGAAGATGACAGCAAGTATGTCAGCAAGTTTCATCCCAAAA 372  
 1249 sIySLeuSerAlaIySAspGluIySLeuIySLeuIySLeuIySLeuIyS 1266  
 376 GAGGAAATATCTGCACCAACAGATGACAGAAATTCGATTCACGATTCG 422  
 1266 LucGlnIleIySAspGlnGlnIySArgCysAspIySLeuThrThr 1282  
 423 ACCAAAGAAATTAAGCATCAGACAGCAACATCTCATTCCTTTAACAA 472  
 1283 IleThrGluLeuGlnIleArgGlyIleGlnIySLeuAlaIySLeuIyS 1299  
 473 ATCAAGAGATTCGAAAGAAATTCAGAAATTCGATTCGATTCGATTCG 522  
 1299 aIySAspAsnIySylIeThrGluLeuAsnAspArgGluIySLeuIyS 1315  
 523 AAAAGTAATTAACATATCACTCACTGCTTATATCATGTGGAAAGCTCGAG 571  
 seq\_name: /SID57/5/cydata/geneseq/geneseqn/NO2001.DMT:AA129101  
 seq\_documentation\_block:  
 ID AA129101 standard; cDNA; 441 BP.  
 AC AA129101;  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Colon tumour related determined cDNA sequence for clone CT2-55.  
 XX  
 KM Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 KM tumorigenic; gene therapy; vaccine; colonic cancer; ss.  
 OS Homo sapiens.  
 XX  
 XX MO200149716-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 XX 29-DEC-2000; 2000MO-US35596.  
 XX  
 XX 30-DEC-1999; 99US-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0645811.  
 XX  
 XX (CORI-) CORIAX CORP.  
 XX  
 PI Xu J., Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jlang Y;  
 XX MPI: 2001-441847/47.  
 XX  
 XX Colon tumor associated proteins and nucleic acids useful for the  
 XX prevention, diagnosis and treatment of colonic cancer -



alignment\_scores:  
 Quality: 674.50 Length: 394  
 Ratio: 2.498 Gaps: 6  
 Percent similarity: 68.528 Percent identity: 38.071

alignment\_block:  
 US-09-750-590-2 x AAH2255 ..

Align seg 1/1 to: AAH2255 from: 1 to: 1434

```

17 AspTPraSnlvSTYrAspAspArIgluMeLArAlaAlaAlaIuArGLyAs 33
   ::
171 GAGTGGACAGAAATGATAGACCGCTACGACGCGCTGAGAAATGAGACA 220
   pVALgluIuYValSerSerIleIleuAlaIleuIleuIleuValAsnProGly 50
   ::
221 TCGCGAAGAGTGGCCCTACTGCTGGCGACAGAGGACCACTGCCACA 270
   50 yALeAspValaIGluIYArGSerAlaPheNIsValaIValaIAserLySGly 66
   ::
271 AACGACAGTGGAGGCGAACGCCGCTTCACATCTGCTGCTCCAGAAAGA 320
   AsnIuGluCYLeuAsnAGAACGCCGCTTCACATCTGCTGCTCCAGAAAGA 320
   ::
321 CAGCTGGAATGCTCCAGGGTCATATTACATGCTGGTGGATGCGAACG 370
   83 rSeAspThraIaGlyArGAsnAlaIleuAlaIleuAlaIleuIleuIleu 100
   ::
371 CCAGAAATCTACCGACACAGCGCTTAAATCTCGACGCCAGAACAGGC 420
   100 lALaIleuCYLeuGluIleuIleuIleuIleuIleuIleuIleuIleu 116
   421 ACCATGAAATGATCAGAGACGCTTCATGCTAAATGCTCAAGCCAAAGT 470
   117 ValAspIleuGluIYArGThraIaIleuIleuAspAlaIleuAlaIAspCY 133
   ::
471 GNGACAGCTGTGGAAAGAACCTTACATTATCGACGCGCTCAGAGGCTG 520
   133 sProSerSerIleGluIleuLeuCYAspHISglYAlaIAserValAsnAla 150
   ::
521 CCTCAAGCTGTGCAGATTCTTCGAAACAGACGCCCATTAACCTCA 570
   150 yAspValaIAspGlyArGThrProLeuValaIleuAlaIleuIleuCYArG 166
   ::
571 AAGATTGGATGGGAATATACCGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
   167 ProThrIleCYsGluIleuLeuIleuAspArgIYAlaAspIleAsnSerAr 183
   621 AGCGATGCTGTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 670
   183 gAspLySGluAsnArGThrAlaIleuMeIleuIleuIleuIleuIleu 200
   ::
671 GACGAAAGTGGACAGACGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCT 720
   200 yAspAlaIValaIGluIleuIleuIleuIleuIleuIleuIleuIleu 216
   721 CTAGCGCTGTGGAGCTTAAATTAAGAGCGACGACCAAGCTGATGTA 770
   217 AspAlaIleuGlyHISAspSerSerTYrTYrAlaAlaAlaIleuIleuAsp 223
   ::
771 GATTCTCTGGATAGACATCGCTCATATTATTCGAAACCTCCAGAAATGCG 820
   223 uAspIleuIleuIleu..... 239
   821 AGCAATTAAGAGCGCTCTTAATTAAGAAATCTCCAGATCGATGATTAA 870
   240 ..... 1YsThraIAser 243
   871 AGACCCCAACAAACCAAGACGATGACGACGCTCTAAATATACCTCA 920
   244 GluAsnSerAluYsGlyArGluIleuTrpLySGlyProSerIleuGI 260

```

```

921 GAAAGAAATGACATCCCAAAAAAGCAAAAGCTCCACGCCCTCTATCG 970
   ::
260 nGluAsnAspIleuSerGluIleuMeIleuAspIleuAlaIleuIleuIleu 277
   ::
971 TCTTACCCAGTTGATGATGCTCTCTCCAGAAATCAATCAATCTGACTG 1020
   277 lArGGLuHISgluAsnIleGluAspIleuGluIleuAsnIleuAspLeu 293
   ::
1021 CAGTATCGGGAAG.....GATCTGATATTTTGTGTCAGAACCCCTG 1064
   294 lIeGluAlaGluIleuIleuIleuIleuIleuIleuIleuIleuIleu 310
   1065 AAGCTGAGATCACTGCTTATACGAGAAAGAGAGAGAGAGAGAGAGAG 1114
   310 sValaAsnGlyIleuGluIleuIleuIleuIleuIleuIleuIleu 327
   ::
1115 TACTACAGGT.....GCTGATA 1131
   327 sPLeu.....GluSerGluIleuIleuIleuIleuIleuIleuIleu 338
   1132 GCTTATTGGATTAATAGTCTCAAGCTGACAGCAAGATCTCTCTCA 1181
   339 lEualaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 352
   ::
1182 TTGCAGCAAAAGTGTCTCTCTTACCTTACCAATATAGAGATTACAGCA 1231
   352 rIleGluAlaIleuIleuIleuIleuIleuIleuIleuIleuIleu 369
   1232 TAAATTACAGGCCCAATACCC...AGAGAGCGGAAACAGCACTTACCT 1278
   369 lYserGlySerHISpHeArGlySGluAspMet 379
   1279 TTGACTATACATCTCCACCAACCTGACTTG 1310
   ::
seq_name: /STD2/9/cgdata/geneseq/geneseq/NA2001.DAT:AA159207
seq_documentation_block:
ID AA159207 standard; cDNA; 1712 bp.
XX
AC AA159207:
XX
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SFO ID NO 1410.
XX
XX Human; nonpolar; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system disease; Huntington's disease; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemorefract;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukemia; ss.
XX
XX Homo sapiens.
XX
XX MO20015312-AL.
XX
XX 26-JUL-2001.
XX
XX
XX 26-DEC-2000; 2000NO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 23-FEB-2000; 2000US-0522112.
XX
XX 09-JUL-2000; 2000US-0522112.
XX
XX 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0654350.
XX
XX 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HISE-) HISEO INC.
XX
XX Tang YT, Liu C, Asundi Y, Chen R, Ma Y, Qian XB, Ren F, Wang D:

```





```

357 LysSerArgPheLeuTyT-PheGluSerAspHisLeuGlySerGlySerHis 373
|||||
360 AANTGGAAATTT.....TCTGAGAGACTTTCACAGGAAATTGGCA 3077
|||||
373 Phe..... 374
3078 TATGGAGGAAATACAGCAAGAACTTAAGATTAATTCACAAAGATG 3127
|||||
375 .....ArgLysGluAspMetLeuLysGlnGlyGlnMetTyT 387
|||||
3128 TTGGCATATGATATAAAAGAGATTTGGAGCTTAATAATTCACAAACATA 3177
|||||
388 MetThrAspSerGlnCysThrSerThiGlyMetLeuProValHisMetGln 404
|||||
3178 ACTGCAATGTTATACGATATATGAG.....ATATATGACCAACA 3215
|||||
404 PArgSerMetLeuArgProLeu.....GluLeuHisLeuProA 417
|||||
3216 AAGCAAGATATTTCTTTATATACAGAGAAATATATCCCAACAAATGT 3285
|||||
417 angLHisSerTySerGluangGluLysGlnLysGlnLysGlnLys 431
|||||
3266 TACAGAGCTGTATACGAAAGAAAGCAACATTAAGCTACGACCTTAAGGAA 3315
|||||
432 .....GluHisMetArgThrPheCysAsn 439
|||||
3316 AATATTTGAATGACCATGTAAGAACCCAGAGATTAAGACTTCTGGGCA 3365
|||||
439 pSerAlaLysGlnAspArgLeuLysLeuGlnAsnGlnLeuHisLeuLys 456
|||||
3366 TGAAGCTTAAGAAAGCAAGAGATATGTTGCACAGAAAGAAACCATGCCA 3415
|||||
456 aLaladLysLysValaLeuHisLeuGlnCysGluAsnValLysGln 471
|||||
3416 TAAAGAAAGAGAGACCTTTCTTAGACCTGTGACAGCTGACAGCTGAG 3465
|||||
472 .....AspSerArgGlnGlnHisLeuArgHisLeuGlnHisArgAl 483
|||||
3466 GAGGAAAACTTAAGGAAAGAGCCAGCACTCCAAACAAACAGCAACA 3515
|||||
483 dLeuLysAspValaGlnLysArgMetTyTArgLysGlnGlyLysValLysG 500
|||||
3516 ACTCTTAATGTACAGAGAGATGAGTGAAGATGCAAGAAACATTAATG 3565
|||||
500 ImeGlnThrHisPheLeuHisLeuLysGlnHisLeuThrSerArgPhe 516
|||||
3566 AA..... 3567
517 AlaThrGlyAsnHisArgLeuMetGlnLysLeuLysAspGlnLeuLysAs 533
|||||
3568 .....ATGAGAGATTTAAAGATTAAGATTAATTAAGAA 3596
|||||
533 pMetLysValLysTyTArgLysLysAsnArgLysGlnLysGlnLysLeu 550
|||||
3597 CAAGAGATTTGACATTCGACATATGCAACAGAGAGCTTGAGTGGCTC 3646
|||||
550 angLHisLeuArgHisLeuMetLeuValaGlnGlnHisPheLysArgPhe 566
|||||
3647 ACAGAACTATATGAAAT.....TATGAGAGAGGAA..... 3678
|||||
567 GlyLysLeuMetGlnGluAsnLysArgLeuGlnLysGlnLysLeuSer 583
|||||
3679 .....TCTATACCAAGAAAGAAAGAACTTCAAGAGAAATTAACAGAGCAT 3725
|||||
583 agLLeuGlnArgGlnLysArgLysArgLysLeuThrGlnMetGlnGly 599
|||||
3726 TGAAGACGAGAGACCACTTAGAGCATATTAAGAGAAATTAAGAGCA 3775
|||||
600 .....GlnLeuLysAspLeuSerAlaLysLeuHisLeuSerTlePro 613
|||||
3776 CAGGCTCTACAAACCAAGAA.....GACATTAAGATTCGCT...CATATTCAC 3819
|||||
614 AlnGlnLysPheGlnLysAsnMetLysSerLeuLeuSerAsnGlnLysAsn 630

```

```

3820 CTAAAGAGACCAAGAAACCTATTGATATGATACCTTAAGAGAGAGCTATCTGA 3869
|||||
630 ctLysAlaLysLysLeuHisAspValaGlnLysGlnLysArgSerLeu 647
|||||
3870 GAGACAGCTCAATTAATTAATTAATCTGAG..... 3897
|||||
647 angLThrArgProLeuLysArgGlnLeuGlnHisLeuLysAlaLysLeu 663
|||||
3898 .....GACTTGAAGAAATTCACATACCAATTTA 3924
|||||
664 AlaGlnHisValLysProGlnLysGlnLysGlnLys..... 676
|||||
3925 CAGAGAGAGATC...CCAGGCTCTATGAGAGAGAGAGAGCTATCTGCTAA 3971
|||||
677 .....SerArgLeuGlnLysSerGlnLysGlnLysGlnLys 4021
|||||
3972 TGTCAAAAAATCATATGATACAGCAAGCAATGATTAAGATCTGATAT 4021
|||||
689 rglLethrGlnLeuThrSerTyAsnGlnThr.....LeuHisLysGln 703
|||||
4022 TACAGAGAGCTCAACCAAGAGAGCTCAACCAAGAGAGAGAGAGAGAGAG 4071
|||||
704 HisGlnLysValaLysLeuAspHisLysLeuLeuThrGln.....Val 718
|||||
4072 ATGAGAGAGCTGACGTGTGTGTGAAATTTCAAGAGAGAGAGAGAGAGAT 4121
|||||
718 HisAsnLeuThrThiGlnMetLysAsnValProLeuLysValaSerGln 735
|||||
4122 AAATCTTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4166
|||||
735 HisMetLysLysSerHisAspValaLysAspValaLysAspValaLys 751
|||||
4169 CCTGTATGTATACATGACAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAT 4218
|||||
752 SerAspValaThrHisLysTyThr.....GlnLysLysLeuGlnMet 766
|||||
4219 GCTAAATCCAGAGAGCTCAAGCAACCAAGAGAGAGAGAGAGAGAGAGAG 4268
|||||
766 LysLeuLeuMetGlnLysAsnHisSerLeuLysArgValaLysSerArg 783
|||||
4269 AGA.....AAGCAATGAACTACCAACCAATCTGATGAGAGATG 4309
|||||
783 LysThrValPheLeuProProGln.....ArgHisGlnLysGln 795
|||||
4310 AGCAA...TTCAACCAAGAGATTCAGCACTATCAATCAATCAATCAATCA 4356
|||||
796 MetHisAlaLeuLysSerAsnHisLeuThrGlnLeuLysArgAlaMet 812
|||||
4357 ATGTGTGATATGTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4406
|||||
812 LysLeuLysLysCysGlnLysArgGlnGlnLysLysLysSerLeuHis 829
|||||
4407 TGTAGCTTAG 4456
|||||
829 erGluAsnAspLeuLysLysLysLysSerHisGlnLysTyValPro... 844
|||||
845 ValLysThrHisGlnGlnLysLysLysLysLysLysLysLysLysLys 854
|||||
4507 CTGAGAACTGAAG 4556
|||||
855 .....LeuSerSerThiLeuAspLysThrAsnArg 865
|||||
4557 AGAGAGAACTTAATGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 4606
|||||
865 LysLeuValaAspValaLysLysLysLysGlnLysGlnLysLysLysLys 877
|||||
4607 AAATATCAACCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4656
|||||
878 .....GlnGluPheValLysHisLysAsn 885

```



[illegible]



```

450  engluLeuAlaHisValAlaGluCysValAlaLeuAlaLeuGluCys 466
1036 .....
467  GluAlaValLysGluAspSerAspGlnGlnIleuLysGlnIleuAlaSpral 483
1037 .....
483  AlaLysAspValGlnLysArgMetCysGlnSerGlnGluCysValLysG 500
1067 .....
500  ImeGlnIleHisPheLeuAlaLeuLysGlnHisIleuHisSerAspAla 516
1113  AGCTGACGTGAGATGTTG.....TCTCATCTC 1118
517  AlaPheGlyAsnHisArgLeu..... 523
1139  GAAACCTCATCTCATAGATCATCTGATGAACTGAAAGGAAAGAGAC 1188
524  MetGlnGlnIleuLysAspGlnIleuLysAspMetLysValLysArgGln 540
1189  TACGTAAAGAGCTCAAGAAAGAAATTTCTGAGGTGAGAGCAAGATGCAC 1238
540  LysAlaSerAlaGluValGlnLysAlaArgGlnGlnIleuLysGlnAspGlu 556
1239  ATCTAAAGCGGAACTTTAGCAAGCTACACAAAGAGAGAGAGAGAGAG 1288
557  .....MetLeuValGlnGluPheLysArgAspGlnLysLysle 569
1289  CAGCATCGGTTTACACCTCAAGTGAATTTAATCAATTACATACAGCAACT 1338
569  ImeGlnGlnLysAlaArgLeuGlnLysGlnLysSerMetCysGlnLysG 586
1339  TCTGACACAGAGCGCAACTAGGGAACTCATGAGTGCAGAGAC 1388
586  LysArgGlnLysArgGlnLysLysLeuGlnLysGlnIleuGlnLysLys 602
1389  ACGAGACGCTTTCAGATGAAAGTTGATGATTAAGAACACAGAGTGGCT 1438
603  AspLeuSerAlaLysLeuAlaLeuSerIleProAlaGlnLysPheGlnAs 619
1439  GATTTACAACTCAAACTT.....TCTCGGTTTCAAGAGCA 1473
619  ImeLysSerLeuLeuSerAsn.....GluLeuAsn.....GluL 631
1474  GTTGAGAGAAAGTTTACAAATTTACAGAAATTCAGCATCAATTAGATA 1523
631  YSALALYS..... 633
1524  AACAACAGCAACAGCTACAGAAACAGAGCTCTCAGCAAAAGCAGCAG 1573
634  ...LysLeuIleAspValGlnLysArgLysGlnLysArgSerLeuAsnGlu 649
1574  GCAAAAGCTTCAGAAAGCTCAGATGATTTGACCAAGAGTCTACGTCAAT 1623
649  rArgProLeuLysArgGlnLysGlnLysLeuLysAlaLysLeuAlaGln 665
1634  TGGCGATTAAGACCAAAAGATCCAGAACTTGATTTTACAGAGAA 1673
666  .....HisValLysProGlnLysGlnLysGlnIleuLysSer 677
1674  GTAAGAAGAAATTTATTTCACTACTAGAAAAAGAAAGAAAGGATTTTAT 1723
678  ArgLeuGlnGlnLysSerGlnLysLeuGlnLysArgGlnIleHisGlnLeu 694
1724  AAAATTCAGCGTGTGAAGAGAGACAGCTGTT.....GTAAACAGATTACA 1770
694  rSerLysAsnGlnIleuGlnLysGlnLysLys.....GluLysV 707
1771  AGAAAAAACCATACATCAACAGAGAGAGATGATCACTAACAGAGAGAC 1820
707  aLysLeuAspAsnLysLeuLeuThrGlnGlnValAsnAsnLeuThr 723
1821  TGAAGATCATCTCAAGAAAGCTTAAGAACAGCCGAGAAATTTGATGAC 1870
724  GluMet.....LysAspValProLeuLysValSerGlnLysLysLys 738
1871  CAGTACAGAGACAGAAAGGACATCTTAAGACCTGCACAAACCGTGGCT 1920
738  sSerHisAspValIleValAspLeuAsnLysLysLysSerAspValT 755
1921  TTCCCTTGAAGAACTAGTGCATGAATGAATTAATGTCATTAATGAAGAA 1970
755  hHisLysTyrThrGlnLysLysLeuGlnLysLys.....LysLeu 768
1971  AGGAAAGGTCTCCAGCTTGACATACAGATTAAGAGCCAAAGCCAAAG 2020
769  LeuMet..... 770
2021  TTACTATCAACAGAGACAGAAAGAACTGCTCAAGAGCTGATCTTCGAA 2070
771  .....GluAsnAla.....SerLeuS 776
2071  TCATTTGACACAGCTCAAAAGCTCAAGATTTACAGATTAACAGACAGAT 2120
776  eLysAsnValSerArgLeuGlnLysValPheIleProProAlaGlnHis 792
2121  ATTAAGATTACTACTAGTGTGATCAGTCACTGCTCAAGTTACAGAGAAC 2170
793  GluLysGlnMetAlaLeuLysSerAsnIleHisGlnLysLysLysG 809
2171  CAGAAACATTCGACAGCTCGAAGATCATCTTAAAGAAATATTAAGAA 2220
809  nLeuSerGlnLeuAsnLysLysCysGlnLysAspGlnGlnLysLysLys 826
2221  ATATCTCTCTTTAGACAGCAAAACCAAGAGCTACAGAGCTCAATTAAGA 2270
826  eLysLeuSerGlnAsnAspLysLysLysLysThrMetSerHisGlnLys 842
2271  AACTAGAAAGCTGAAAGCTTGAAGATTAAGAACAGCAAGAGAGCTTTG 2320
843  ValProValLysThrHisGlnGlnLysThrAlaLeuSerSerThrLe 859
2321  CAGATCTTACACAGCAAAAGACAGCTGACAGATTTAGAGCTCAGAGC 2370
859  uAspLysThrAsnArgLysLeu..... 866
2371  CAGCAAGATTTGAGAACACCTTGAAATTAAGAGAAATAGTACAGTA 2420
867  ....ValAspValLysLys.....CysGlnAspIleAsnGln 878
2421  CAGATGATGATCTACAGAAAAAATGTGAAGCCCTTGAAAGTATACAGAA 2470
879  GluPheValLysLysAspGlnAsnGlnIleuLysLysArgAsnLysG 895
2471  AAGCTTACAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGTTTGA 2520
895  uAsnThrGlnAsnGlnValLysAlaGlnLysLysLysLysLysLysG 912
2521  AACTTTAAGTCAAGAAACAAAGATTCAGATGAGAAATTTACAGAA 2570
912  LysGln..... 913
2571  TTCAACAGACATACAGACATCAAAAGATTAAGTGAAGAAAGAGACT 2620
914  .....LysMetSerGlnLysArgLysSerMetLysLysValGlnAs 927
2621  TTAAATACAGACCTTTCTACAGTAAAGAGAAACATCAAAAGCTTTCTGA 2670
927  PAsnSerAlaGlnIleuAlaLysTyrLysLysSerGlnGlnGln..... 942
2671  TTCTTTGAAAGAACTTAAAGATTTTGAAGAGAGAAAGCAATTAAGAGAA 2720
943  .....HisValThrLeuHisGlnGlnLysAlaAlaGlnLysArgGln 956

```

```

1165 11etelysAIaserleuAATGGGluysGluGluGluSerGlnInsuTys.. 1180
1166 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1167 ATACTAAAGACCGAGGTGATAAAAGCAAGAGGAGGACCTGGAAGAAAGATT 3702
1168 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1169 1181 11etGluIuValSerLysLysLeuGlnInsuGluIleGlnAsnThrLysGlnA 1197
1170 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1171 1197 TATTGGAAGAAAGACTGATATGATGATGATGATGATGATGATGATGATG 3752
1172 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1173 1197 11etGluLysLysLeuGlnThrATGAGGluIuValValAspLeuSerLysTyrTys 1213
1174 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1175 3773 GATTGAAGACGCTACGAAAGAAATGAG..... 3778
1176 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1177 1214 AlaThrLysSerAspLeuGluThrGlnIleSerAspLeuAsnGluLysLe 1230
1178 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1179 3779 .....GCTAAATCTACCATGCTGATATGATGATGATGATGATGATGAT 3819
1180 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1181 1230 uAlAsnLeuAsnThrLysTyrGluGluIuValCysGluGluIuValLeuHis 1247
1182 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1183 3830 AGCAGCTGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3869
1184 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1185 1247 11etLysLysLysGluIuSerLysLysLysAspGluIuLysGluLeuIleHis 1263
1186 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1187 3870 TTGAAGAAAGACCGAGGTGATAAAAGCAAGAGGAGGAGGAGGAGGAGGAG 3919
1188 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1189 1264 SerIleGluGlnGluIuLysAspGln.....GluArgCys.. 1276
1190 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1191 3930 AGGTTGAGATATATCATGATGAAAGAGAGGAGGAGGAGGAGGAGGAGGAG 3969
1192 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1193 1277 .....AspLysSerLeuThrThrIleThrGluLeuGln 1287
1194 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1195 3970 TAAAGGAAAGGAGGTGATAAAAGCAAGAGGAGGAGGAGGAGGAGGAGGAG 4019
1196 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1197 1287 11etArgTyr.....IleGlnGlnSerLysLysLeu 1296
1198 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1199 4020 AAAGAAAGCTGGATAATATCATGAGCTGAGCAGTGTGAGGAGGAGGAGGAG 4059
1200 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1201 1297 11etGluIuLysAspAsnLysIleThrGluLeuLeuAsn 1309
1202 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1203 4070 AACCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4108
1204 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1205 seq_name: /SIDS2/gcgdata/geneseq/geneseq/VN2001.DAT:AA160039
1206 seq_documentation_block:
1207 ID AA160039 standard; CDNA: 8527 BP.
1208 XX
1209 AC AA160039:
1210 XX
1211 XX 22-OCT-2001 (first entry)
1212 XX
1213 DE Human polynucleotide seq ID no 4028.
1214 XX
1215 XX Human: nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
1216 XX peripheral nervous system; neuropathy; central nervous system; MS;
1217 XX Alzheimer's; Parkinson's disease; huntington; lymphoma; chemotheric;
1218 XX amyotrophic lateral sclerosis;
1219 XX leukemias; ss.
1220 XX
1221 OS Homo sapiens.
1222 XX
1223 PN W020015312-A1.
1224 XX
1225 PD 26-JUL-2001.
1226 XX
1227 PR 26-DEC-2000; 2000MO-US$4263.
1228 XX
1229 PR 21-JAN-2000; 2000US-0488725.
1230 PR 25-APR-2000; 2000US-0552317.
1231 PR 09-JUL-2000; 2000US-0558042.
1232 PR 19-JUL-2000; 2000US-0620312.
1233 PR 03-AUG-2000; 2000US-065210.
1234 PR 14-SEP-2000; 2000US-0662191.
1235 PR

```





Seq. name : /SID2/gsgdata/geneseq/geneseq/MA2000.DAT:AAA09326

seq\_documentation\_block:

ID ID AAA09326 standard; DNA: 6773 BP.

XX AA  
AA AAA09326:  
DT 10-AUG-2000 (first entry)  
DE Human cancer associated antigen precursor DNA, clone NT-RBN-53.  
XX DE  
XX renal cancer; cancer associated antigen precursor; diagnosis:  
XX cytosolic; KIM0316 mRNA tag; ss.  
XX OS  
XX Homo sapiens.  
XX PN  
PN MO20000587-A2.  
PD 13-APR-2000.  
XX PD  
XX 04-OCT-1999; 99MO-USZ2873.  
XX PR  
PR 05-OCT-1998; 98US-0166300.  
XX PR  
PR 05-OCT-1998; 98US-0166350.  
XX XX  
XX (LUDM-) LUDMW INST CANCER RES.  
XX PA  
PA Ohtani Y, Gout I, Tureci O, Sahin U, Pfeundschn M, Scanlan MJ, Stockert E, Chen Y, Old LJ, Jager E, Knuth A;  
XX DR  
DR WPI: 2000-303774/26.  
PT Preventing, diagnosing and/or treating disorders associated with  
XX abnormal expression of human cancer associated antigens  
XX  
XX Claim 57; Page 97-99; 121pp; English.  
XX  
XX AAA09321-45 were isolated by SREX screening from a renal cancer  
XX cell line 1973/20/4. Homology searching revealed that the

428 LYNCH, BARNETT, AND KATZ



1967 ATGAAGTCTCTCTGAGACAAAGAGTATTGTCAAGCTGAGTCAAGTC 2016  
727 nval.....ProLeuValSerGluGluMetLysLys...SerH 740  
2017 TCTTATGAGGAAAACAACTAACTAGTCTGAAAAAAGCAAGTGAAGA 2066  
740 LAspValLLeuValAspAspLeuAspLysLysLeuSerAspValThrHis 756  
2067 GGGATTGGAGGTT...TTTTCTCTCAAAA...GAAGTTTCTCTT 2110  
757 Lys.....TyrThrGluLysLysLeuMetGluLysLeuLeu 769  
2111 AAGAACATATTACTCAATTAGAAAAGAACTACCTA.....ATGCT 2154  
769 uMetGluAsnAlaSerLeuSerLysAsnVal.....SerArgLeuGluTr 784  
2155 TGAAGACCAAGATATTAAATTAACCTGCTCAAAATGACAGATGACA 2204  
784 hrValPheLeuProGluArgHisGlu.....LysGluMetPheAla 798  
2205 AGTATTCTTAAACTAGTGTATGCTTTCTTAAAGAAATGGATCA 2254  
799 LeuLysSerAsnLLeuThrGluLeuLys..... 807  
2255 GAAGTTTCAGAGACGCTGAGAGAAAGAGCTGTTAAATGCTTCAAGC 2304  
808 .....LysGluLeuSerGluLeuAsn...LysLysCysGly 818  
2305 AGCTGCGAACTCTTGCGAAAAATTAATGAGAAAATTCACACCTGGCTT 2354  
819 ..GluAspGlnGluLysLLeuYrSerLeuMetSerGluAsnAspLeu 834  
2355 TTGACCTGCTGAAAAGATATTAGAGTTAGAAAAGAGATTTAAAGTCTT 2404  
835 LysLysrMetSerHisGlnTyrValProValLysThr..... 847  
2405 CAAGAGAGAGCTAGCTGACGTGAGAGACTTAAGCTTTATTGAGAGA 2454  
848 HisGluGluLLeuLysThrAlaLeuSerSerThrLeuAspLysThrAsn 864  
2455 CTTCAGACAGAGAAAGTCTCTTA.....AGCA 2483  
864 rGluLeuValAspValLysLysCysGluAspLLeuAsnGluPhe 880  
2484 AAGAGTTAGAGAAATACAGTACAGAAAAGCCCTGACGCTGATCTT 2533  
881 ValLysLLeuAspGluAsnGluLLeuLysLysArgAsnLeuGluAsnH 897  
2534 CTGAAAATGAAAGATGCTCAATGAA.....AAAGACAGCTTCAAAATCA 2577  
897 rGluAsnGluValLysAlaGluLysLysSer.....LeuA 909  
2578 GAATCTTTAATTCAAGTGAAGAGATCTCAAACTAGTACCAAAAGTC 2627  
909 rGluHisGluGluLysMetSerGluLeuAspLysSerMetLysLysVal 925  
2628 AAATGCTAAATGAAAA.....GAAAAATGCTTTTAAAGGAA 2665  
926 GluAspAsnSerAlaGluLLeuLeuAlaGlnLysAspGluLeuAspTr 959  
2666 CATGAAAACCTAAAGCCACTACTA.....GAACAAAAAG 2700  
942 uLLeuValThrPheHisGluGluLLeuAlaGlnLysAspGluLeuAspTr 959  
2701 ATTACGAGATAGAGACAGAGAGCTGATACATATTAAAG.....GATT 2741  
959 hrLLeuGlnLysLysLLeuLysLeuLysThrAlaProLLeuLLeuGlu 975  
2742 CCTTACCAAAATGACCTTCTGAAAAAATGATGCTGCTGCTCAAGTAAA 2791  
976 GluLysGluLysrPheLysLLeuThrGluLysGluLysLysGluLLeu 992  
2792 GAGTGGAGAGAAATTAAGAAATCTGCAAAAAAGATGCAAGAA..... 2836  
992 uSerGlnGlnThrGluLysrPheThrSerGluGluGluAlaLysLysC 1009  
2837 .....AGGAGAGAAAATTAATTAAGCTTAATTAATTAATTAATTAAG 1115  
1009 yLysGlnGluAsnAspLysLeuLysLysGluLLeuLysrPheGlnLys 1025  
2880 CAAGAAAAGAACTAGATTCCACAGAAAAGAGCCAGAGCTGAGAGAA 2929  
1026 AspLeuLysAspLysAsnValHisLLeuAsnSerTrGluThrGluTr 1042  
2930 GAAGTTGAA..... 2938  
1042 gAlaLeuSerArgLysrThrGluGluLeuAsnAspGluLLeuLysArgLeu 1059  
2939 ..TCTCTTCATGACAAAAGAGACAGATTATCTGCTTCATGACATCTCA 2987  
1059 euGln.....LysrYrThrGluAlaLysLysGlu 1068  
2988 TTCAAGAGACAGAAAGCTATTAAAGATCTTTATTAAGATTAAGAGCG 3037  
1069 LysGluLysLeu...ValGlnGlu.....AsnAlaLysLeuThr 1080  
3038 TCAGACCACTGATGATGGGAAAAAGACGCTGATTAATTTTGAGCATCG 3087  
1080 rSerGluLLeuAlaAlaGlnThrLeuLeuGlnLysGluHisValProL 1097  
3088 TATTGAAGACCTTACAGAACAA.....TTAAAGAAATTCGACTTCAAGT 3131  
1097 euGlnGlnValGluSerLeuLysSerLeuSerGluLysrThrLLeuThr 1113  
3132 GTGAAACAAATTAATTCTGATTAATGAAGATCTCTGCTGATTTAGAGCA 3181  
1114 LeuLysGluGluLeuLysrThrLysGlnArgLysrYrGluLysGlnGlu 1130  
3182 TTACAG..... 3187  
1130 nThrValThrGlnLeuArgGlnMetLeuGlnLeuGlnLysAsnSerSerV 1147  
3188 .....TCATATCCAAATTTATTAAGAG 3209  
1147 alProLeuAlaGlnHisLeuGlnValLysGluLysLLeuLysGluVal 1163  
3210 TACAGATTTTAAAGATCCAGAGAACCCAGAACCAATGCTAGACAAAGAAATTA 3259  
1164 GlyLLeuLLeuLysLLeuAspLysrGluLysGluGluLysrGlnAsnLys 1180  
3260 GAACTGAAAACCTTCAGAAAAGACAGAGATTAAGAGAACTGCTCAAGC 3309  
1180 rThrGluGluLLeuLysrLysLLeuGlnSerGluLLeuLysAsnThrLysGlnA 1197  
3310 TGTAAATGAACTTGAGAACTTCAAGCTGACAACTGAACTGAAAGAAAGAAAC 3359  
1197 LaLeuLysLysLeuGlnThrArgLysValValAspLeuSerLysrYrTr 1213  
3360 AGCTTCAGAA.....ACATGCAAGATTAAGTACCTGCTTAAAGAGAT 3403  
1214 AlaThrLys.....SerAspLeuGlnLysrThrGlu 1222  
3404 GCCCAACAAACCAATTCATGATGAAATATGAAATAGCTGATATTAGAGCTTT 3453  
1222 nLLeuSerAspLeuAsnGluLysrLeuAlaAsnLeuAsnArgLysrYrGlu 1239  
3454 GATGAAAGAACTAAATGAAAGCTTACATTAATTAAGAGAAAGCAATGAGAA 3503  
1239 LuValCysGlnGluLysLLeuHisAlaLysLysLysGluLysSerAlaLys 1255  
3504 ATTGGAGCAAGAAATTAATTAATTCAAAAAGACAAAGAACAAACCTTCACAA 3553  
1256 AspGluLysGluLeuLeuHisPheSerLLeuGluL..... 1267  
3554 CAAGAAATTAATCTTAAATGATCTTCAAGTACAGCAATATGAGAAAAAAA 3603

1268 ..... Glut1leuYAFSPG 1272  
 3604 CACCAAAATCAACGACATTCCTGTGMAAACCAAAAGCAACGACGCGAGATT 3633  
 1272 IINGINGIUAFCYASAPYSSerLeuThrIleThgIuGInuIaNG 1288  
 .....  
 3654 CAAGACAGCAAGAAATCATCATCTTAATVACTTCACACCTCTTAAAGCT 3705  
 1289 AAGIleGInuIaSerAlaYgGIIIGIUAIAIbYASAPASuYgIIleTh 1303  
 .....  
 3704 GACGTGACGCAACCCACGACAGACTGAAAGCTTAAATATACAGCTCC 3753  
 1305 FCuIeuleuAmaSPVALGUAryLeuYgGInaIaIeuaInuGlyeUS 1322  
 .....  
 3754 TGAATVACATCAAGCAACAAATTCACAGCAACCTGAAA..... 3796  
 1332 eGInuIeuthrTYrGlySeGlySerProSerYAFaGInuSerGInu 1338  
 ..... ACCTGCGGAACAGCAACGACGCTACG 3823  
 3797 .....  
 1339 ILeaSPSerLeuGInuIaIuValaGySerLeuGInuGInuIaIaIa 1355  
 .....  
 3824 CTAACTCATTCACGACGAGAGTGCACGACGCTCAAGAGAGGCGCCCTGC 3873  
 1355 PALaSPaTgGInuISGInuIuValIleAlaIeYrAGThrIIS 1370  
 .....  
 3874 TGCAGACGCAAGCAACGCTCACTCACTCAATTCGAAATTCGAAACCTTAC 3919  
 .....  
 seq\_name: /SID52/gcgdata/geneseq/gcneseq/MA1998.DAT:AAV21518  
 seq\_documentation\_block:  
 ID AAV21518 standard: cDNA: 5661 BP.  
 AC AAV21518:  
 XX 17-ANG-1998 (first entry)  
 XX  
 DE Rattus norvegicus mutant alpha-myosin heavy chain transgene.  
 XX  
 KM Actin binding domain: alpha myosin heavy chain; bridging; mouse;  
 KM transgenic mammal; congestive heart failure; study; treatment; diet;  
 KM exercised; effects; identification; hypertrophic cardio-myopathy;  
 KM dilated or hypertrophic cardiomyopathy; acute aortic regurgitation;  
 KM tricuspid stenosis; constrictive pericarditis; hypertension;  
 KM aortic regurgitation; aortic stenosis; aortic valve disease;  
 KM aortic regurgitation; aortic stenosis; aortic valve disease;  
 KM hypertrophic cardiomyopathy; aortic stenosis; aortic valve disease;  
 KM Paget's disease; transgene: ss.  
 XX  
 OS Rattus norvegicus.  
 OS Synthetic.  
 PH  
 FH Key Location/Qualifiers  
 FT 1:5661  
 FT CDS /feature B  
 FT /product= "mutant alpha-myosin heavy chain"  
 PN W09813476-A1.  
 PN  
 PD 02-APR-1998.  
 PE 26-SEP-1997. 97MO-US17296.  
 XX  
 XX 26-SEP-1996. 96US-0026742.  
 XX  
 PA (LEIN/) LEINMAND LA.  
 PI Vlkstrom KI.  
 DR WPI: 1998-230690/20.  
 DR P-BSDB: AAM54241.  
 XX  
 PT Transgenic mouse models for congestive heart failure and

PT hypertrophic cardio-myopathy - used to study molecular and cellular  
 PT events, identify potential therapeutic agents, assess effects of  
 PT diet etc.  
 XX  
 XX  
 XX Claim 39: Pages 43-53; 75pp: English.  
 XX  
 PS The sequence is that of a mutant rat alpha-myosin heavy chain  
 CC gene which was used in the development of transgenic mammals,  
 CC specifically mice. They can be used as a model for studying congestive  
 CC heart failure (CHF) or hypertrophic cardiomyopathy. Such animals  
 CC are used to study molecular and cellular events associated with  
 CC CHF. To identify compounds for treating CHF, and in evaluating  
 CC effects of diet, and exercise on CHF. Conditions associated with CHF  
 CC that can be evaluated this way are dilated or hypertrophic  
 CC cardiomyopathy, acute aortic regurgitation, endocarditis, ischemic  
 CC heart disease, hypertension, primary myocardial diseases, valvular  
 CC or pericardial disease: hypertrophic disease; anemia; arteriovenous  
 CC fistula; heart failure and Paget's disease.  
 XX  
 SO Sequence 5661 BP, 1572 A; 1446 C; 1797 G; 846 T; 0 other:

alignment\_scores:  
 Quality: 600.00 Length: 1370  
 Ratio: 0.805 Gaps: 54  
 Percent similarity: 54.380 Percent identity: 20.949

alignment\_block:  
 US-09-750-590-2 x AAV21518 ..

Align seg 1/1 to: AAV21518 from: 1 to: 5661

237 ThrIeuleuYrThrAlaSerGlu..... AnSerAaInuYsG1 249  
 1699 ACACCTCTTCCACGATCTCTGTGATACCGGTCGACAGTGGAAAG 1748  
 .....  
 249 YArGluIeUTrPLySGLYProSerLeuGIn..... 260  
 .....  
 1749 CAAGAGAGCAAGAAAGGCTCATCTCTCCAGCAAGCTCTCTCTCTCC 1798  
 .....  
 261 .. GInaTASuIeUSerGInuIeUleuASpGIIuValaInThrIIS..... 274  
 1799 ACCGGAAATCTGAACAGCTGATGCAAACTGAGACACCACTCT 1848  
 .....  
 275 ..... SerASInaTgGInuISuIaInIleG1 284  
 1849 CACTTTCCTGCTGATATCCCAATGAGCGGAGAGCTCCAGGGGAT 1898  
 .....  
 284 nASpLeuGInuIleGInuASpLeuYs..... GIna 296  
 1899 GAGCAACCCCTGTGTATGCAACAGCTGATGCAACGAGAGCTGGAG 1948  
 .....  
 296 rGLeuArY..... LyS 299  
 .....  
 1949 GATATCCCATCTGACAGAGGCTTCCCAACGACATCTTATGAGGAGC 1998  
 .....  
 300 IIGInuIaGInuIaYrIleU..... 307  
 .....  
 1999 TTCCGGACAGSTATGCAATCTGACACCAACCACTCTGAGAGCCA 2048  
 .....  
 308 ... LeuASpLyValaInuGInuIeUleuASpGIIuValaIM 323  
 .....  
 2049 ATTCAATGATAGCGGAGAGGCGCT..... GAGAACTCC 2083  
 .....  
 333 eValaIaIaSPaSPLeuGInuSerGInuYgGInuYsIeU..... 335  
 2084 TGGGCTCTCCGACATGACCAACCACTCAACAAATTTGGCGACCAACAG 2133  
 .....  
 336 ..... LySerLeuIeUAlaIaIuYgIIIGIUAIAIbYASAPASuYgIIleGInuI 348  
 .....  
 2134 GTTCTTCAAGAGGGGCTCTGTGGGGCTGCGAGACAGATGCAAGTGA 2183

```

348 useLeuArgThrIle.....Glu 355
    ||| |||
2184 GAGCTGACCCCATCATCAACGATTCACGACGACGACGACG 2233
355 IalaLeuysSerArgPheIys.....Phe 361
    ||| ||| |||
2234 TCAATGCACTTGATTCACAGAAAGATGCTGACGACGACGCTCTG 2283
361 ..... 361
2284 GTTATCCAGTGGACATCCGCCCTTCATAGGGGGTCAAGAAATGGCCCTG 2333
362 .....TyrPheGluSerAspHisLeuGlySerGlySerHisPheA 375
    ||| ||| ||| |||
2334 GATGACACCTCTACTCTAAAGATCAAGCCGCTCTGAAAGGCGACAGACG 2383
375 TgIysGluAspMetLeuLeuIysGluGlyGlnMetIysMetHisPheSer 391
    ||| ||| ||| |||
2384 ACAGAGGATGGCCAAATCAAGAAAGAGGATGGCGGACGATCAAGATGCA 2433
392 GlnCysThrSerThrGlyMetProValHisMetGlnSerArg.....Se 406
    ||| ||| ||| |||
2434 CTAGAGAGATGTATAGGCTCGCCGACAGAGAGCTGAGAGAAATGTGCTC 2483
406 PheLeuAlaMetArgProLeu...GluLeuAlaLeuProAsnGlnAlaSerTyrS 422
    ||| ||| ||| |||
2484 CCTGCTCGAGAGAAAGAAATGACCTCGACGCTCAAGCTCAAGCTAGAGCAAAAG 2533
422 ergIu.....AsnGluIleLeuIysGlu 430
    ||| ||| ||| |||
2534 ACACACCTGGACATGCGAGAGACGCTGGGACACCACTGATCAAGACAAAG 2583
431 LeuGluAlaMetLArgThrPheCysAspSerAlaIys.....G1 443
2584 ATCCACGTG.....GAGGCCAAGCTGAAAGACATGAC 2615
443 nAspArgLeuIysLeuGlnAsnGluLeu...AlaHisIysValAlaGlnC 459
    ||| ||| ||| |||
2616 CGAGAGCTCGAGAGACGAGAGAGATGAAAGCCGACCTACGCGCCANA 2665
459 yAlaIysAlaLeuAlaLeuGluCysGluArgValIysGluAspSerAspGlu 475
    ||| ||| ||| |||
2666 AAGCAAGCTGAGAACACAGCTCAAGACTCAAGAAAGATGTGATGATC 2715
476 GlnIleIysGlnLeuGluAspAlaLeuIysArgValIleGlnIleAspMetIys 492
    ||| ||| ||| |||
2716 .....CTGAGCTGACCTGGCCAAAGCTGAGAAAGAAAGCA 2753
492 rGluSerGlnIleIysValIysGlnMetGlnThrHisPheLeuAlaLeu 509
    ||| ||| ||| |||
2754 GCGACACAGACAGACAGATTAAGACCTGACAGAGAGATGGCCGAGTGG 2803
509 ySGIuHisLeuThrSerAspAlaLeuThrGlnAsnIleAspMetIysGlu 525
    ||| ||| ||| |||
2804 ACGAGATCATTCGCCAAG.....CTGACCAAGAGAGAAAGACTTTCA 2847
526 GluLeuIysAspGlnLeuIysAsp..... 533
    ||| ||| ||| |||
2848 GAGGCCACACGACGAAAGCCCTAGATGACCTTACAGTGAAGACAAAGGT 2897
534 .....MetIysValIys.....T 288
2898 CAACACATGACCAACAGCTTAAAGTCAACCTGAGACGACAGAGGATGATC 2947
538 yTgIuIysAlaSerAlaGluValGlyIysLeuArgAsnGlnIleIys... 553
    ||| ||| ||| |||
2948 TGGAGGATTCCTCGAGACAGAGAAAGATGGCTGATGACCTGAGCA 2997
553 ..... 553
2998 GCAAAGCAAGCTGAGAGGTGACATGACGACGACGACGAGACATCAT 3047
554 .....GlnAsnGlu.....MetLeuValGlnIlePheIysArgAspG 566
3048 GAGCTGAGAGACACAAAGCTTCAAGCTGAGAGAGAAAGCTTAAAGAAAG 3097
    ||| ||| ||| |||
566 IuGlyIysLeuMetGluIuAsnIysArgLeuGluIuIysGluLeuSerMet 582
    ||| ||| ||| |||
3098 AGTTTGACATCATGTCAGCAACAGTAAATTAAGAGACAGACGACGCTG 3147
583 CysGluLeuGluArgGluIysArgGlyArgIysLeuThrGlnMetGluI 599
    ||| ||| ||| |||
3148 GGC.....CTTCAAGCTGACAAAG.....AACTCAAGAGAAACCAAGGC 3185
599 yGlnLeuIysAspLeuSerAlaIysLeuAlaLeuSerIleProAlaGln 615
    ||| ||| ||| |||
3186 ACCCATCGAGAGACTGAGAGAGAGAGCTGAAAGCGGACGACACCGGG 3235
616 ..LysPheGluAsnMetIysSerLeuLeuSerAsnGluLeuAsnGluIys 631
    ||| ||| ||| |||
3236 CCAAGCTGACAAAGCTGCTGCTCAACCTCAACCGGACGCTGAGAGATC 3285
632 AlaIysIleu.....IleAspValGlnArg 640
3286 ACTGAGAGCTTAAAGAAAGCCGCTGGGCGACATCTGTGCAATTAAGAT 3335
640 gGluIysArgIysSerLeuAsnGluThrArgProLeuIysArgGluLeuG 657
    ||| ||| ||| |||
3336 GAAACAGAAAGCGCGAGGCC...GAGTTCCAGAAAGTATGCGCGGACCTGG 3382
657 IuAsnLeuIysAlaIysLeuAlaGlnHisValIysProGluIuHisGlu 673
    ||| ||| ||| |||
3383 AG.....CAAGCCACGCTG.....CACATGAG 3405
674 GlnLeuIysSerArgLeuGluIuIysSer.....GlyIuLe 686
    ||| ||| ||| |||
3406 GCCACAGCCGCGGCCCTGCGAAAGAGACGCGACGACGCTGGCGGAGCT 3455
686 uGlyIysArgIleThrGluLeuThrSerIysAsnGlnThrLeuGluIysG 703
    ||| ||| ||| |||
3456 GCGCGACAGATAGACATCTCAAGCGGCTGATGAGAGAGCTGAGAAAG 3505
703 IuIleGluIysValCysLeuAspAsnIysLeuLeuThrGlnIleValAsn 719
    ||| ||| ||| |||
3506 AAGAGAGCGATTCAAAGT.....GAGCTGGAT 3534
720 AsnLeuThrThrGluMetIysAsnValProLeuIysValSerGluIuMe 736
    ||| ||| ||| |||
3535 GAGCTCACTGTCAATGAGACATC.....ATCCAGCCCAAGACCAACT 3581
736 IlyIysSerHisAspValIleValAspAsnIleIys.....L 750
    ||| ||| ||| |||
3582 GAGAGAAAGTCCCGACACTGTGAGACGACAGAGAAAGATACCGCGTGA 3631
750 yAlaSerAspValThrHisIysTyrThrGluIysIysGluIuMetGlu 766
    ||| ||| ||| |||
3632 AGTTTGAAAGAGACAGACGCTCTCAATGATCTCAACACACAGACGCC 3681
767 IysLeuLeuMetGluAsnAlaSerLeuSerIleIysAsnValSerArgIuLe 783
    ||| ||| ||| |||
3682 AAGCTCAGACAGACATGCGCACTGTGCTGCGCAATCGAAGAAAGGA 3731
783 uThrValPheIleProGluArgHisGluIysGluIuMetMetAlaLeu 800
    ||| ||| ||| |||
3732 GGCACGATTTGGACGCTACCCGGGCAAG.....CTCTCTCTATA 3772
800 ySerAsnIleThrGluIysGlnIleIysGlnIleIysSerGluLeuAsnIys 816
    ||| ||| ||| |||
3773 CCGAGCAGATGAGACCTCAAGAGACGACGCTGAGAGAGAAAGCCAG... 819
817 CysGlyIuAspGlnGluIysIleTyrSerLeuMetSerGluAsnAsn 833
    ||| ||| ||| |||
3820 .....GCCAAGAAATGCTTGGCCCAACGATCTCAAGTACGCGGATGA 3863
833 pLeuIysIysThrMetSerHisGlnIysValProValIysThrHisGln 850
    ||| ||| ||| |||

```

[illegible]

seq\_name: /SID2/seqdata/geneseq/geneseq/NAL1996.DAT.AAT34291

seq\_documentation\_block:

ID AAT34291 standard: cDNA to mRNA: 6175 BP.

AC AAT34291:

XX 25-MAR-1997 (first entry)

DE Coding sequence for smooth muscle myosin heavy chain SM1.

XX Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;

KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; restenosis;

KW associated adenovirus; coronary artery catheterisation; sclerotic artery;

ss.

XX Mus musculus.

XX Key Location/Qualifiers

XX CDS 105..6023

XX /tag= a

XX /product= SM1

XX MO9623069-A1.

XX 01-AUG-1996.

XX 25-JAN-1996: 96MO-JP00134.

XX 25-JAN-1995: 95JP-0010085.

XX (OSAP) OSAKA PREFECTURAL GOVERNMENT.

XX (VBS-) VESSEL RES LAB CO LTD.

XX Sugawara M, Takahashi K;

XX WPI: 1996-362693/36.

XX P-Psdb: AAM00024.

XX DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in

XX a vector for gene therapy for reduction of re-stenosis following

XX coronary artery catheterisation

XX Claim 5: Page 14-27; 42pp: Japanese.

XX This sequence represents the coding sequence for the smooth muscle myosin

XX heavy chain SM1 isoform protein. This SM1 coding sequence was isolated

XX from a mouse embryo cDNA library using probes based on the rabbit smooth

XX muscle myosin heavy chain SM2 isoform. This sequence is included in the

XX gene therapy vector of the invention. The vector of the invention

XX consists of this sequence inserted into a retrovirus, adenovirus,

XX associated adenovirus or animal-expression plasmid vector (such as pCMX2

XX or pCMX208). The vector can be used in the gene therapy treatment of

XX restenosis, particularly for the reduction of restenosis occurring

XX following coronary artery catheterisation for widening of sclerotic

XX arteries.

XX Sequence 6175 BP: 1846 A; 1455 C; 1856 G; 1018 T; 0 other:

alignment\_scores:

Quality: 595.00

Ratio: 50.887

Percent Similarity: 21.865

Percent Identity: 21.865

Length: 1578

Gaps: 72

Align seg 1/1 to: AAT34291 from: 1 to: 6175

4 CysTrpPheSerCysAla.....ProLysAsnArgGlnAlaAlaAspTr 18

1705 TGTGTGATGAGATGCTGTGTCCCAAGACGTACAGAAAGCTTTGTG 1754

18 pAsnLysTrpAspArgMetLeuAlaGluArgGlyAspVal 34

1755 GAGAACCTAT.....GCTCAGACAGCGAATA 1783

35 GluIysValSerSerLeuAlaLysGlyValAsnProGlyLysLe 51

1784 CCCCAATTCCAGAACCCACAGACGTAAAGACAAAGC..... 1822

51 uAspValGlnGlyArgSerLeuPheHisValAlaSerLysGlyAsn 68

1823 .....AAGTTTCATCTCATCTCACTATGCTGGGAAGGTGAC 1859

68 euGIuCyS...LeuAsnAlaIleLeuIleHisGly..... 78

1860 TACAAATCAAGTCCGTGCTGACACCAAGAACTGACCCCTTAATGACA 1909

79 ValAspIleThr..... 83

1910 TGTGACATCATCTCATCTCCCTCTGACAGTTTGTGCTGACCTGT 1959

84 .....SerAspThrAlaGlyArgAsnAla 92

1960 GGAGAGATGAGACCGCATTTGTGGGCTGACACAGATGCGCAAGTACT 2009

92 euHisLeuAlaAlaLysTrpGlyHisLeuLysGlyLeuGlnLeu 108

2010 GAGACTCAGTCCGACG.....TGCCTC...AAGACCA 2041

109 GlnTrpAsnCySProThrGlnHisValAspLeuGln.....GlyArg 122

2042 AAGGCGATGTCGCGACCGCTGGACAGCTCTCAACAAAGACAGTGGGA 2091

122 gThr.....AlaLeuHis 126

2092 AACATGATGCTACACCTGCGCATACACAGCTTAATCTTGCTGCTGATC 2141

127 .....AspAlaIleMetAlaIleAspCysArgProSerSerIle 137

2142 ATCCCAACCATGAGAAAGCTCTGCGACAGCTGATGATTCCTATGAGCT 2191

138 GlnLeuLysCysAspHisGlyAlaSerValAsnAlaLysAspAla 154

2192 .....GDNACAGCTGCTGACACAGCTGTCTT...GG 2220

154 YArgThrProLeuVal.....LeuAlaThrGlnMetCysArgPro 168

2221 AAGCATCCGATGCTCTGCGACAGGCTCTCCCAACAGATGTCTTCAA 2270

168 hTrIeCySLeuLeuIleAspArgLys..... 177

2271 GAGTTCCGACAGCGATGAGATCTCTGCGACAGACGCAATCCCAAGC 2320

178 .....AlaAspIleAsnSerArgLysGlnAsnArgThr 189

2321 CTTGATGATGAGAAAGCAAGCGTATCTCATGATCAAGACCTCGAAC 2370

189 r.....AlaLeuMetLeuG 194

2371 TTGACCTTAACCTGTACAGATTGGCGAGCAAAATCTCTCCGACGC 2420

194 ly.....Cys 195

2421 GGGGTCTGTGCCCCACTGACAGCAAGCAACACCTAAATTAATCTGATGT 2470

196 GluTrpGlyCysLysAspAlaValGlnValLeuIle..... 207

2471 CATCATGCGCTTCCAGGCAATGCTGTCTGCTGCTGCTGCTGCAAGCCCT 2520

208 .....LysAsnGlyAlaAspValThrLeuLeuAspAlaLeuGlyHisAsp 222

2521 TCACCAAGAGCGACACAGCTGTGACACCAAGGATGATTCAGAGAAC 2570



```

4255 .....
4256 ACATGACATTCACGCTCTCCAGACTTAAAGAAAGAGCTTCGCG 4255
743 .....
4256 AACGACATCGAGGCTCATCGAGAGAGGAGAGAGGCTTACAGAAAGGA 4255
751 euserSerPvalThrluYstYrThGlulYstLeuGluMetGluYs 767
4306 TGGAGGCTTCACGCTCCAGCTACGAGAGAGAGGCGGCTGCTATGAGAA 4355
768 LeuMetGluAuaLaserSerLysAsnValSerPgluLeuLTh 784
4356 ..... CTGGAGAAACCAAGAACGGCTC..... 4379
784 rValPheIleProGluAuaGhLsGluYsGluMetMetAlaLeuYs 801
4380 ..... 4404
801 eAsnIleThrGluLeuYsLysIleSerGluLeuAsnYstYs 817
4405 TGGCTTGGACAAACCGAGGCACTGTATCCATCTGGAAAGAG... 4451
818 GlyAlaSerGlnLysIleYrSerLysMetSerGluAuaAsnAprle 834
4452 ..... 4481
834 uYstYrThMetSerLysGlnYrValProValYstThHsGluLut 851
4482 .GAGAGAGCTCTCCAGATGCG..... 4508
851 lElyrThAlaLeuSerThrluAuaPrYrThAsnAtGluLeuVal 867
4509 ..... GATGAGAGAGACCGACCTGAGCA 4532
868 AsPvalYstYsCysGluAspIleAsnGlnLupheValYstIleY 884
4533 GAGGCGAGAGAAAG... GAGCAAAAGCTTTCTCTACGCGGCGCT 4579
884 AsPGLuaAndLulleuYrAsnLeuGluAsnThGlnAsnLuv 901
4580 GAGAGAAACCTCGAGACCAAGAACGCTGGAGAGACCAAGATGC 4629
901 aLysAlaGluYrTlaserLeuAtrGlnLsGluYstMetSerGly 917
4630 TCGAACCTGAG..... ATGGAGAC 4649
918 LeuYstSerMetYrYrValGlnAsnAsnSerAlaLutIleLeuAl 934
4650 CTGGTCACCTCAAGATGATGTAGCAAGAACTGGCTAGACG... 4694
934 aLysYrLysSerGlnLsGluLlLevalYrThrLeuAsnGluLl 951
4695 ..... GAGAACTCAAGCGCTGTGGAGACCAAGATGAAAGAA 4737
951 lAlaLutLysAlaGluLeu... AsPThrIleGlnLsCylleYrLeu 966
4738 AAACCGCTGGAGAGCTGAGAGATCACTGAGCGGCAATGAGATGC 4787
967 YstYrAlaProIleLeuSerLeuGlu..... GluYstGluYr 979
4788 AAGCTGGGCTAAGAGTCAATGCAATGCAAGGCGCTTGGTAAG 4837
979 GluYrLeuYstAlaThrlu..... LysGluLeuYstGluLutLeu 993
4838 CGATCTCCAGCTCGGATGAGCAAGATGAGAGAGAGAGAGAGCTTC 4887
993 eGlnGlnThGlnLysYrAsnThSerGlnLutAlaLysYsCys 1009
4888 ACCGAGCGCTCACAGATATGAGAGAACTGAGAAATGAGAGAGAGAG 4937
1010 LysGlnLutAsnAprYrLeuYstYsGluLlLeuThrThrluYs 1026

```

```

4938 AAGCTCTGGCGGCGGAGCTTAAAGAGAG..... CTGGAGAGGGA 4978
1026 PleuYstAsPvalAsnValHsIleGluAuaSerYrGluThrluAua 1043
4979 CCTAAAGACCTTACGCTCCAGGCTGACCTGAGCTTACAGCAAGAGAGAG 5028
1043 lAlaSerAtrGluYstThGluLutLeuAuaAtrGlnLutLeuYstPleuLeu 1059
5029 AAGCATTCAGAGCTTCGAGAGATTCAGAGCTGCTGATGAGAAATTCGA 5078
1060 GlnYrYrThrluAlaYs..... 1066
5079 AAGAGCTGATGATGCGGCTCCGACGAGATGAGATCTTGGCACCTC 5128
1067 ..... LysGluYstGluYsLeuValGluLutAsnAlaYstGln 1080
5129 AAAAGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5178
1080 hTrSerGluLlLeuValAlaGlnThrluLeuLutLysGlnHsValPro 1096
5179 TCCAAAGAGAGCTGCGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 5225
1097 LeuGluGln..... ValGluSerLeuYstYrSerLysSerGlyTh 1110
5226 CTGGAG 5275
1110 rIleGluThrluYstGluLutLeuYstYrThrluAuaYrGlyrGluL 1127
5276 G... AATPACCTGAGATGAG..... AAGCGCGGCTCG... 5306
1127 YstGluGlnThrluValThrluGlnLeuAuaGlnMetLeuGluAsnGln 1143
5307 ..... GAGGCAAGATGCGGCACTGAGAGAGAGAGAGAGAGAGAG 5351
1144 AsnSerSerValProLeuAlaGlnHsLeuGln..... ValY 1156
5352 GCGGAG 5401
1156 sGluAlaPheLutYstGluValGlyLlLeuYstAlaSerLeuYrGluL 1173
5402 TGGCAAGCTGAGATGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAG 5451
1173 YstGluGluYstGlnAsnYstThrluGluValSerYstLeuGlnSer 1189
5452 ATGGAG 5501
1190 GlnLlLeuAuaPrYrThrluYstYrLysGluLutLysGluLutAla 1206
5502 AAGTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5531
1206 lValAsnLeuSerYrYrValAlaThrluYstSerAsnLeuGluThrlu 1223
5532 ..... GCGAGCTCAAGCTCATCTTGGCTGCGCTGCGCTGAGCA 5571
1223 lAserPleuAuaGlnLutYstLeuAlaLeuYstAsnValYrGluGlu 1239
5572 TGGAGAGCTGAGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 5618
1240 ValCysGluGluValLeuYstAlaYstYrGluYstLeuSerAlaYs 1256
5619 ..... GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5647
1256 rGluYstYs..... GluLeuYstAlaSerSerIleGlnGlnLut 1270
5648 GAG 5697
1270 YAsnGln..... GlnGluYrCysAsnYrSerLeuYrThrluThrlu 1285
5698 CAG 5747
1286 LeuGlnAtrGatGllLeuGlnSerAlaYstGln..... lLeuGlu 1299
5748 CTGGAG 5797

```

```

1299 alysaaspasnylsilethrcileuleuleuaspval.....G 1312
    ::: ::||::: :: ||:::|
5798 CAACCGACGAGAGCTCAGCGGAGCTAGATGAGCCACAGAGACAAATG 5847
    :::::|
1312 LARGleuLYSCInAlaLeuAnglyLeu...SerGInLeuThrTyrcly 1327
    :::::|
5848 AGGCGATGGCGCCGTGAGGTGAGAGGCGCTCAGAGACCAACACTGAGAGAGGA 5897
    :::::|
1328 SerGlySer.....ProSerLYsArg 1334
    :::::|
5898 AAGAGGCTTCATTGTTCCTTCAGAGAG 5927
    :::::|

```





```

286 TTTCAGCTTGACAGCTCGACAGCTGGCCGCCGAGAGCAACACAGACAGTTC 335
289 GUAAR ..... 290
336 GAAATGCAAAAAACCCAGGCTCAACAGCTGAAAAAGGAGAAATCAAAATTC 385
291 ..... GluAspLeuIysGluArgLeuArgLysIleGlnGlnG 303
386 GATGCAAAATATGCTAAAGCTGACAAACACTGACAGAAAGAAATCTCTTATC 435
303 LuchAqIleLeuLeuAspLysValAsnLysLeuLysGlnLysLeuAsn 319
436 AACTTCAGCTCAAGAGCTCAACAGCTCAATTCACGAGAGGACAGCTGAAAT 485
320 GUGUUAUAlMetValAlaAspAspLeuGluSerGluLysGlnLysLeuLys 336
486 TCAGCGCAAAAAACAAATAGAAAAACCTGCAACGACAACTTAAAACTGTGATC 435
336 BserLeuLeuLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 353
536 ATCTGAGCTTGAAGAACGACAAACAGCTGCGAGAGCTGCGAAGTCTGCTTC 585
353 IeclLysLysLeuLysSerArgPheLysTyrPheGlnLysSer ..... 365
586 TGAATCCATGCAATACCCACAAAAATTTTACACACTGCTCAACACCA 635
366 AspHISLeuLysSerLysSerLysPhe ..... 374
636 AGTCATATATATATGCTCTTCCAAAGTATGAAAGTATGAAAAATATTA 685
374 ..... 374
686 TAAAGAGGTGAAGACGAAAAAGATTAGACGACAGATTAAAGCTTCGC 735
375 ..... ArgLysGluAspMetLeuLysGlnLysGlnLysMetLysMetThr 389
735 ACGCTAAAGAAACAGACAGACAGCTCTTCAAGAGAGAAACAGCAATATACCGC 785
390 Asp ..... SerGlnCysThrSerThr ..... 396
786 GACATTGCCCGCATCAGCTTCATCATCTGCTTCATGACGACGACAGA 835
397 GlyMetProValAlaMetGlnSerArgSerMetLeuArgProLeuGluI 413
836 CAAAGCCCAAGATCATCTTCATCATCTTCATCAACAGCAAGCTCAATTAAG 865
413 euAlaLeuProAsnGlnAlaSerTyrSerGlnAsnGlnLysLeuLys 429
886 GAGATTTC ..... TCATCATCTTACTTTTCGCGGACACTA ..... 920
430 GluLeuGluAlaMetArgThrPheCysAspSerAlaLysGln ..... 930
921 GAGGTGACCTCAAGTGATCATCACTTGCATTAAGGAGAAAGACAGCTAA 970
971 TAGCACTTCTTGGCAATTCAGACAGCTCATCTCATCTTGGATCAATTA 1020
444 ..... AspArgLeuLys 447
447 LysLeuGlnAsn ..... GluLeuAlaLysValAlaGluLys ..... 459
1021 AAGCGAGATTCAGAGCTTAAGAACAGATTAAATGATTGAGACTAGCC 1070
460 ..... LysAla 461
1071 CTCGAAAGACATGAAAAAGAAATGAAAGCGCAAGCTGAATATCTTCAGA 1120
461 AlaLeuAlaLeuLysGlnLysGlnLysGlnLys ..... 472
1121 ACTCCAACTCCAACTCGAGAAAGCAAAAGTGGCAATTAATTGTAAGAGAGA 1170
473 ..... SerArgGlnLysLys ..... 478

```

```

1171 AAGTTTGACAAATGATAGACATGATCATGATGAGCAACAGCAATATC 1220
479 ..... GlnLeuGluAspAlaLeuLysAspVal 487
1221 GACACGCGTCAACCAAGTATCTGATCTGATGAAACAAACCTGAAAAATTT 1270
487 L ..... GlnLysArgMetLysPheLys ..... 494
1271 GACGCAAGATTTGATGATGCTGACGACAAACATGCAAAAGGCGCAATGTT 1320
495 ..... GluLysValLysGlnMetGlnThrLysPhe ..... 505
1321 CTCGTGACACGAAATTTAAGCAAAAAAGAAAGAGATTTCAAGAGGACGTC 1370
506 ..... LeuAlaLeu 508
1371 TCCGCTCAACAGCTGTTCTTCCAAACAGCTGACACAGCAAGCTCCAAAT 1420
508 LysGlnLysLysLysSerAsp ..... AlaLeuThrGlnAsnIleArgLys 523
1421 GAAGCGCAACCTCACCACGAGATTCACGAAAGCAAGAAATATGCAACAG 1470
523 euMetGlnLysLeuLysAspGlnLysLysAspMetLysValLysTyrGlu 539
1471 TCTGTGAGCTGAACTGATATATCTCACTGATATACAGCAAGCTAGCA 1520
540 Gly ..... AlaSerAlaGlnLysGlnLys 547
1521 AACAAATTGCAAGATTAAAGCAAAAGCTGTGCAAGCTGACAGAGCGTT 1570
547 sLeuAlaArgAsnGlnLysGlnLysAsnGlnMet ..... LeuValGlnLysP 562
1571 CCAGCGATGTCATCAATCAAGAGAAATAGAGCTGACAGAAAGCATGAGAGAA 1620
562 IleLysArgAspGlnLysLysLysMetGlnLysLysLysArgLysGlnLys 578
1621 TCAAGAAAGAAACAACTCTTTAAGATGACCTCGAG ..... CAAAG 1664
579 GluLeuSerMetLysGlnLysGlnLysArgLysArgLys ..... 591
1665 GCCAAGAAAGTCTGCCACTGTGAGCAAGACTCAAGAAATCATCAACAGTC 714
592 ..... ArgLysLeuThrG 596
1715 TTTAAATCAAGCGAGAAATTTTGCAGAAAGAAATGAAACCGCAAAATTCCT 1764
596 LysMetGlnLysGlnLeuLysAspLeuSerAlaLysLeuAlaLysSerIle 612
1765 CTCAGCAAGCAAGCTTAAAGATCTTCAAGAAATTA ..... 1802
613 ProAlaGlnLysPheGlnLysMetLysSerLeuLeuSerAsnGlnLys 629
1803 ..... AsnArgCysAspMetLysCysThr ..... 1829
629 nGlnLysAlaLysLysLeuIle ..... AspValGlnLysArgLysGlnLys 645
1830 GAAAACTGAAAGCTTCTGTGCTGATCTGAAAGAGACGCAAGTTGTT 1878
645 IleLeuAsnGlnThrArgProLeuLysArgLysGlnLysLeuLysAla 661
1879 CTCAGAGCTTTGAAAGAAAGAGACATCACTGATCAACATCTGAAT 1928
662 LysLeu ..... AlaGlnIle 666
1929 AACTTAAGCAAGACAGAAAGAGAGCTCAAGAGCTCTGAGCTGTATTA 1978
666 sValLysProGlnLysGlnLysGlnLysSerArgLeuGlnLysLys 683
1979 GTTAAAGAAAGAAATTTGCAAAATTTGAA ..... GAGAGAGAA 2019
683 et ..... GlyGlnLysLysArgLysLysLysLysLys 693
2020 CTCGTCTTCTTCTTGAAAGAGTGAAGAAAGAAACCTTTAACTGATG 2069

```



```

3768 GAAATTTCTGATAGTATATGCTCCAGCTGATTCACATTAAGAGCTAT 3817
1169 rleuaygclulysgcluglulserglnanlysrthrgcluglulvalserL 1186
3818 GCTACAGAAATAGAGAAATTAACCTTCAGAAAGAGACAGAGAGAGT 3867
1186 yaleuInsergclulieglInanrthlye.....Glnala 1197
3868 GCCTGCACCATCAATTACACAAATTCAGAGAGCTTCGAACACGCAAT 3917
1198 leuylayleuInluthrAgclulvalvalapleu..... 1209
3918 TTGCACAGACATGTCAGACAGAAATTAAGTGCCTTAAGACCTGAAAT 3967
1210 .....serlyrthlyalrthlyserAapleuInrthrgclul 1223
3968 ACATGCGCAAGAAAGTATATTCAGGCGCTTCAGTCTTCACAAAGTC 4017
1223 leserAap.....leuanclulysleualaan 1232
4018 AAMACGCAATGCAACCTTCAGTCTTCGCAACACATGATGAAAG 4067
1233 leuanrlyrlyrthrgcluglulvalcygclulvalleuInlaleuylsly 1249
4068 GCTAT.....GACCTACAGAAAGATATG.....GAAATCTGCAGAGCTGAA 4111
1249 slaygclulserleaylsergclulyscluldeuInleuInlshpserleG 1266
4112 GATTCAGACTGCTAACT..... 4127
1266 lucInglulleylserpInglInglulnarlycyasrlyserleuThrthr 1282
4128 .....GACCTAAATGATTCAGAGCTGAAAGT..... 4154
1283 lIerhrcululserglnanrAagcluglulserleaylserleuInleuInl 1299
4155 ATCCAGCCACTAGCAAAATTCGCAAGAGCTGAGGAAA..... 4193
1299 alYAsrAanlyslerthrgcluldeuInanrlyalgluInleuylsG 1316
4194 .....CTACTAAATGCAAGTT.....A 4209
1316 lnaaleuInsergclulsergcluldeuThrlyrly..... 1327
4210 AATATTTAATGATGACAGCTGCTTCCTTCATGCTGATGATTAAGCAAGC 4259
1328 .....se 1328
4260 ATACAGAGAGTGATTTGGTGCACAAACCAATGAACAGCAGCTCTGTG 4309
1328 rclYserp.....sepl 1333
4310 TTTCGCTCATTTGGACAGAGTAATTCCTACAGAGCCTTCATCATTTGAG 4359
1333 yAATgclInsergcluldeu.....leAsrserleuInglInglulvalser 1348
4360 ACAAGAGAACTCAATTCGACTTTCGCAATTCGACAGAAATTCCTATCT 4409
1349 leuInglIngluldeuInleAsrAanrlyrthlysergclulvalleuInl 1365
4410 TTTCAGAGTGAACCAAAATTTACATGCTTCAGAGCTTCAGAGT..... 4454
1365 alIeYrAtrthlyleuInserleaylaleglInglInlylmeleAap..... 1380
4455 .....ACCTCAAAATGCTCAGAGCTGCAAGCTGATGTGCTCAT 4494
1381 .....GluAsrlyalalale 1385
4495 TAAAGCCGAAATTTGCTTTCGACAACTTCGAGAACTTCAGAGT 4544
1386 AlaleuInleuInlIeInl.....MetArGclInglInlylcyse 1400

```

---

```

4545 GACTTGGTGAAGAGATGATCAAGCTGAGCTTGAGAGAGGAGCTGTTCATC 4594
1400 F 1400
4595 C 4595

seq_name: /cgn2.6/prodata2/1na/PC/TUS.COMB.seq: PCT-US95-16216-2
seq_documentation_block:
; Sequence 2, Application: PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; EXPRESSED PROTEIN: Expressed Kinocytore Protein, and Methods of Use
; CORRESPONDING ADDRESSES: 4
; ADDRESS: Deon Dorfman, Herrell and Skillman
; STREET: 1601 Market Street, Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC DOS 5.0
; SOFTWARE: Patentio Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09/08/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed Janet E.
; REGISTRATION NUMBER: 36,252
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14136 base pairs
; TYPE: nucleic acid double
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-2

alignment_scores:
Quality: 561.00 length: 1567
Ratio: 0.758 gaps: 72
Percent similarity: 47.224 Percent identity: 21.889

alignment_block:
US-09-750-590-2 x PCT-US95-16216-2 ..
Align seg 1/1 to: PCT-US95-16216-2 from: 1 to: 10136
251 Glnuethrlyrlyslsly.....ProserleuInglInl...ArAsy 263
186 GAAAGATGGAAGAGAGGCTGCGCCACCAACAACTTCGAAATTCAGAA 235
263 nleuSerGlnmeleAapclulvalAanrthlyserAanclInrthrgcluln 280
236 GCTTGAAGAGAGCTTCGACAAAGTGAAGAGGAAAGAGCAAGCAAGCT 285
280 lsgIn.....AanlIeInlAsrleuInl 288

```

```

286 TTCACCTTGACAGTCTCGAGGCTGCGGCCCGAGAAACACACAGCAAGTT 335
289 Gtuan..... 290
336 GAAATATGAAAAACCGAGGTACAACTGAAAGGAGAAATCAAAAT 385
291 ..... Gluspleuysgluatrgleuatrglytleuglinc 303
386 GATGCAAAATATGTGAATCTGAGACACACTACAGCAAGAGATTTCTCATG 435
303 lugaltrglietleuleuaspysvalasnglyleuadlueuglnleuan 319
436 AACTTCAGCTCAGAGGTACCAAGTGAATTTCCAGCAAGACACACTGAT 485
320 GluGlualmetvalAlasapleuigluserdlyugsglyleu 336
486 TCAGGCAAAAAACAAATAGAAAACTGACACAGCACTTAAAGGTGTA 535
336 sserleuleualalalyesglulysglulnslgluserleuagthet 353
536 ATCTGACCTTGAAAGAACCCACAAAGCTGCGAGTCTCGAGATGCTC 585
353 leglualleuysersertrpheytytyr-phegluser..... 365
586 TGATTCATGCAATACACCAAAATTTTTCACACCTCCACATCAACCA 635
366 Asphlisleuglyserdlyserthlaph..... 374
636 AGTCAAATTTATGCTGTTCCAAATGATGAAATCTAAAGAAATATA 685
374 ..... 374
686 TAAAGCTTGAGAACCAAAAAAGATTAGCCGACAGCTTAAAGCTTGC 735
375 ..... Arglysgluaspleuleuylsglulysglulmetlyrmetr 389
736 AGCGTAAAAAGCAACCCAGCTCTTCACACAGCCACCATGAAATCCGC 785
380 AHP..... Serdlynglyserthr..... 396
786 GACATTCGCCCGCATACAGCTTCATCATCTGTGTTCTCATGCGACAGA 835
397 . GlymetProvalHismetglinserArgsermetleuanrProleu 413
836 GAAGACCCAGCATCTTTCATCTCAATTCCTCAAAACATCCCAATAGGA 885
413 eualeuleuproanglnalasetrysergluanobliuleuyls 429
886 GAGATTC.....TCATCTCTACTTTTCTGGAACTA..... 920
430 Glueuadlualmeterythrphecysaspselalyegln..... 443
921 GAGGTGACTCAGTGCATCACTTGCAATGCGAAAGAGATGTA 970
444 ..... 970
971 TAGCAGTCTTTCGCAATTCAGACGCTCATCTTTTGATGTCATTA 1020
447 yslengluasn..... Gluleualahlslygvalabglucys..... 459
1021 AACGCGAATTCAGAGCTTAAGAAACAGATTAACGTGTGACACTGCC 1070
460 ..... Lysal 461
1071 CTCGACAGCATGAAAAAGAAATGAAAGCCAGTGAATTAAGTTCAACA 1120
461 Alenalalaleuclucysgluatrgvallysgluasr..... 472
1121 ACTCCAACTCACTCGAAGAAACAGAAAGTGCAATTAATCAAAAAAGCA 1170
473 ..... seraspglucglulleyys..... 478
1171 AAGTTTGAACAATGTAGGATGACTAGTGAAGAAACACACCAATAC 1220
479 ..... Glueuadlualaspleuleuysaspya 487
1221 GACACGCGCTCAACACAGTATACATGCAATTCGACAAACAACTGABAAAT 1270
487 L..... Glnlysartrmetryrcluser..... 494
1271 GACGAGACATTTGATGTCTCACCACAAAATGCAAAAGTGCAGATGTT 1320
495 ..... Gluelylusvalylsglmetcglnthrlspher..... 505
1321 CTCGTGACACAAATTAAGAGAAAGAAAGAGATTCACAGAGAGCTC 1370
506 ..... LeuAlale 508
1371 TCCGTCACACAGCTTCTTCACAAACAGTGCAGAGAGATGATCCAGT 1420
508 ullysgluahlsleuthrserasp..... Alaseralgluvalolyl 547
1421 GAAAGCCACACCTCCACAGGACCTTACACGACCAACCAAGATGATG 1470
523 euwetglulgluleuylasrpleuleuylasrpleuylvalylrcl 539
1471 TCTGCGAGCTGAGCTGATTAAGTCACTGATTAAGCAACAGCTAGAA 1520
540 Gly..... Alaseralgluvalolyl 547
1521 AACATTTTGCAGAGATTAAACGAAAGTTTGACAGCTGAAACAGCTT 1570
547 sleuafgsandlnlleyglulnanglunet..... Leuvalglulup 562
1571 CCAGCGAGTCAATCAAGAGATGATGACGTGAGAGCAACATGAGCAAA 1620
562 helysarfpaspulncllylsleumetcluglunsnlylsarpleugl 578
1621 TGAGAGAGAAAGAACACCTCTCTTAAGATGCTGTGAG..... CAAGAG 1664
579 Glueusermetcylglueuadlualargllylsarargly..... 591
1665 GCCAGAAAGTCTGCCACCTGGAGGAGAACTGAAAGCAACCAAGTG 1714
592 ..... Arglylsleuthrg 596
1715 TTAAATTCAGACGAGATTTTGACAGAAAGAAATGAAAGCAAAATCT 1764
596 lumeretglulgluleuylasrpleuseralylsleualaleuserlle 612
1765 CTCAGGAACCATGTTAAGAGTCTTCAGAAAAATA..... 1802
613 ProalaglylsrpheglunsmetlyserleuleuserAanglueuas 629
1803 ..... AATTCAGCAAGAAACCTCTTGACATTA..... 1829
629 ncllysalalyslylsleulle..... Aspyalgluarglulrtyrgluars 645
1830 GAAAACTGAACCTTCGTGCTGATCTGAGAAAGACAGCAAGATGTT 1878
645 etleuasnglurthrarProleuylsartrglueuglunleuylasr 661
1879 CTCACACCTTTTGAGAAAGAAAGAACATCACTGAAACACTTAATAT 1928
662 lylseu..... Alaclnlnl 666
1929 AAGTTAAACAAGACAGCAAGAAAGATCCAAAGCTTGCTGATGATTTAGA 1978
666 svalylsproglnlglulnslglulnleuylsersartrgluadlucly 683
1979 GTTAAAAAGAAAGAAATATGAAGATGAAA..... GAGAGAAAAA 2019
683 et..... Glyleuendlylylsartrglueuglue 693
2020 CTCGTGTTTCTGTGGAAAGAGTGAAGAAACAAATCTTAACCTCATG 2069

```

[illegible]

```

3768 GAAATTTCTCTAGATAGTTATATCCGAGTGTGCTCATTTACAGCAT 3817
1169 TleuATgclulysglulglulserglulnaslnlythrglulglulserl 1186
||||| :||||| :||||| :||||| :
3818 GCTNAGAGAAATNAGCATTTAAACTTCAGAAAGTGAAGAGAGAGCAT 3867
1186 yslenglnserclulieglulnaslnlythrglulglulserl 1197
||||| :||||| :||||| :||||| :GlnAla 1197
3868 GCCTCCAGCATGATTTACAGCATTTAGAGAGCATTTGAAACCCAGCAT 3917
1198 TleuLysylsleuGlulthraTgclulvalylasplu..... 1209
||||| :||||| :||||| :||||| :
3918 TTCCAAACACACCTGACGACCAAGAAATTTAGCTCTTAAGACTGTGAAT 3967
1210 .....SetlyTylsAlaThrlyserAspLeuLthrlnl 1223
||||| :||||| :||||| :||||| :
3968 AGATCCGAAAGAAAGTATATTTCAGGCCCTCATGATTTGACAGATC 4017
1223 leserAsp.....LeuasnclulysleuAlaasn 1232
||||| :||||| :||||| :||||| :
4018 AANAGCAATGACACACTTCAGTCTCTGCAAAACATCATGACAAAG 4067
1233 LeuasnArgLysTylrclulglulvalylsleuGlulvalleuAlaLyslyl 1249
||||| :||||| :||||| :||||| :
4068 CTGAT.....GAGCTAGAGAAATNTGT.....GAAATCTGAGGCTGACAAA 4111
1249 slsYsluLeuSerAlaLysAspGlulysgluldeuHlsPheSerileg 1266
4112 GTATGAACCTCTACT..... 4127
1266 lnglnclulylsAspGlnlnglnLysArgLysAspLysSerleuThr 1282
4128 .....GAGCTATATGTCAGAGCTGACAGATGT..... 4154
1283 lIethrcluleuGlinsArgylieglulserAlaLysglulieglul 1299
||||| :||||| :||||| :||||| :
4155 ATCCAGCACACTAGCAAAATTCGCAAGAGACCTGAGGAAA..... 4193
1299 alysAspAsnLyslIethrcluleuLeuAspValylglulArgLyslysg 1316
4194 .....CTGCTAGATGAGAT.....A 4209
1316 lAlaLeuasnGlyLeuSerGlnLeuThryrGly..... 1327
4210 AAATATTAAATGATGACGTGCTTCTTCATGCGAGTTAGTACAGAC 4259
1328 .....Se 1328
4260 ATACAGACAGTGAATTTGTGACACCAATGACAGACCTGTGTG 4309
1328 rGlySerPro.....Se 1333
4310 TTGTGCTCATGAGACAGAGTATCTCCACGACACTTGACATGTGAC 4359
1333 ysaTgclulserGlnleu.....lIeasPserleuGlnlnglnlvalylasr 1348
||||| :||||| :||||| :||||| :
4360 ACNAGAGATTCATGAATGACACTCTCCGCAAGCAAGAAATCTTACT 4409
1349 LeuGlnlnglnldeuAlaAspAlaAspArgTglnlnslnglnlvalylleal 1365
||||| :||||| :||||| :||||| :
4410 TTACAAAGTGAACAAATTTTACATGATCAGCAGCTGACGATG..... 4454
1365 alIeTylrGThrlsleuLeuSerAlaAlaGlnlYHlsMetAsp..... 1380
4455 .....AGCTCTAAATGTCAAGCTGACGACCTGTGTGATCTAT 4494
1381 .....GluAspValGlnAla 1385
4495 TAAAGCCGAAATTTGCTTGTGACACCAATCTGACAAATCTTCAAGT 4544
1386 AlAlaLeuGlnlIlelIeGln.....MeLArgGlnlYLeuValLysse 1400
||||| :||||| :||||| :||||| :
4545 GACTGTGTAAGAGAGATGACGTGAGCTGCTGACAGAGGAGCTGCTCATC 4594

```

---

```

seq.name: /cgn2-6/ptodata/2/line/SA_COWB.seq.us-08-328-254-5
1400 F 1400
4595 C 4595
seq.documentatn_block:
Sequence 5, Application us/08328254
GENERAL INFORMATION:
PATIENT NO.: 571002201
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hua
TITLE OF INVENTION: A No. 571002201 Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: CA 92037A
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PATENT IN Release #1.0, Version #1.25
CURRENT PATENT DATA:
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REFERENCE NUMBER: 81,815
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8789 base pairs
TYPE: nucleic acid
STRANDNESS: single
TOPOLOGY: linear
FEATURE:
MOLECULE TYPE: DNA (genomic)
NAME/KEY: CDS
LOCATION: 544..7990
US-08-328-254-5
Alignment_scores:
Quality: 540.00 Length: 1623
Ratio: 0.667 Gaps: 73
Percent Similarity: 49.866 Percent Identity: 21.688
Alignment_block:
US-09-750-590-2 x US-08-328-254-5 ..
Align seg 1/1 to: US-08-328-254-5 from: 1 to: 8789
33 AspValGlnLysValSerSerleu.....AlaLysylsGlnlvalAsnPr 48
||||| :||||| :||||| :||||| :
2551 GACTTAGAGAAATATGTGCAATACTGACGCTGAAAGTATGAGATCGT 2600
48 ogLysLysAspValGlnLysArgLysSerAlaPheNlLysValAlaSerL 65
2601 AACTGACGTGAT.....GATTCAGAGTCA..... 2625
65 yslGlyAsnLeuGlnLysLeuAsnAlaIle..... 74

```





```

586 GluATGluLysArgGluYArgLysLeuThrgLumelGluGluLeuLys 602
4201 CATCCACGAGAGTTACTCCAGAGAGTGAAGAACTTCCAGAGCCCTCATATTC 4250
602 SASPLEuserAlaLysLeuAlaLeuSerIleProAlaGluLysPheGluA 619
4251 TCATTTTGGAAATGCATGCATTAATCTACAGCGAGAACTTTGGAAAGTA 4300
619 smELysSerLeuLeuSerAsnGluLeuAsnGluLysAlaLys...Lys 634
4301 ATGTGACC...AAGCGAATGACACGCTGGAGACAGACAGA 4335
635 LeuIleAspAlaGluArgGluLysIleuLysSerLeuAsnGluThrArgP 651
4336 TTTCCTTATGTGAAAATTCAGCTACGATGAGATCGAGAAAGGCTACG 4385
651 OleuLysArgGluLeuGluLysLeuLysAlaLysGluAlaGlnHisAlaL 668
4386 CATTGACCATGAAACCCCTTACCTGAGAGCTGACTA... 4422
668 yAFroGluGluHisGluLeuLysSerArgLeuGluGluLysSerGly 684
4423 ...GAGGTACTTCAACAGACAGAGCTATGTTACAAA 4458
685 GluLeuGluLysArgIleThrgLumelThrSerLysAsnGluIleLeuGlu 701
4459 ...GACAAATGAAATATAGCA 4475
701 nLysGluIleGluLysValCysLeuAspAsnLysLeu...LeuThrg 716
4476 GAGGCTATT...GTCTGCTGTGAAGAGAACTCTCAGTGTGCAGAA 4519
716 IngIleValAsnAsnLeuThrThrgLumelLysAsnAlaProLeuLys... 731
4520 GTGAGAGAAACCACTCTCGAGATTAATCTATGTGTAAAAAAC 4559
732 ...ValSerGluLumelLysSerHisAspValI 743
4570 ACCGCACTGATCATCGTGTCTGAAAATAAGAACGAGAA... 4608
743 eValAspAspLeuAsnLysLysLeuSerLysValThrIleLysIleThrg 760
4609 ACACAGAGCTTGACTCATCAAACTGACGTCTCCATCTCAGTCATCAG 4657
760 LuLysLysLeuGluLumelGluLysLeuLumelGluLysAsnLysSer 776
4658 TGGCAGAGCCAGAGGAGAGAGAAAGCAGAACTCTTCAAGCTTGTCC 4707
777 LysAsnValSerArgLeuGluLysValPheIleProProGluLysHis... 792
4708 TCTGATGTACGTAGACTG...TTAAAAGAAACACTCT 4745
793 ...GluLysGluLumelLysLeuLysSerA 802
4746 CCAGAGAAAGATCGAGACTTGTGAAAAGACTCCAGGACACGCTTTGA 4795
802 snIleThrgLumelLysGluLysLeuSerGluLeuAsnLys... 815
4796 CAATATGTACGTGAGAAACCAATTCACAACTTAAATATAGAAAGAA 4845
815 ...TTGCTGTCAAGAAATCTGAAGCTGCAAGCCACAGTCAATCAG 4895
816 ...TTATGAAAAGCTGAATGTCTCCAGAGCTTGAAGCCGCACTGTGTGAGA 4945
817 ySGlyGluLysPGLuGluLysIleThrgSerLeuLeuSerGluLysAsnAsp 833
4946 AAGGTGATGTGCATTGAGAGCTGACCTCAACAGAGAGAACTCATCAG 4995

834 LeuLysLysThrMetSerHisGlnIleValProValLysThrHisGlu 850
4996 CTCGAAAGAGCATCGAAAGAACTGAAAGTTCGATTCAGTACGCACTGAAA 5045
850 uLleLysThrAlaLeuSerSerThrLeuAspLysThrAsnArgGluLeu 867
5046 GAACGACCTGCATCATCCAGAAAGAACTGCAAGAAAGCCAGCGAGATG 5095
867 AlaAspValLysLysCysGluAspLe... 876
5096 ATTCTCTTAAAGATTAAGTTCAGAAACCTTGAAAGGAATTCGACATGTCA 5145
877 ...AsnGluLysPheValLysLysAspGluAsn... 887
5146 GAAGAAACCGAGACTGTGATCTCTTGATCCGAGAAATTCGAAAGAGAA 5195
888 ...GluIleLeuLysArgAsnLeuLysAsnThrHisAsnGluValLysA 903
5196 ACTGAAGACTTGAAGAACAAATATAGAAAGAGCCGCAAGCTGAAAG 5245
903 Le...GluThrIleSerLeuAspGluLysGluLysLysMetSer 916
5246 TTTTGATTTAGACTTGTGCTGATTAAGTGTGAAAGAAATCTGAGCA 5295
917 GlyLeuArgLysSerMetLysLysValGluAspAsnSerLysGlu... 931
5296 ...AAACAAATACAAAGAAACAAAGAGCTGTGTC 5327
932 ...IleLeuAlaLysIleLys...Lys 939
5328 AGAATACAGCAAGTTACTCTCTTCAATTTAAAGTCTGTGAGAAAGAG 5377
939 eGrgGluGluLysIleValThrLeuHisGluGluLysAlaIleAlaGluLysArg 955
5378 ACCAAGCAAGAGAT...CAATCAAGAAAGAA...TCAAAAGCT 5415
956 GluLeuAspThrIleGluGluCysIleLysLeuLysIleAlaProIle 972
5416 GCAGTGSAGATGCTTCAGAAATGATTAAGAGACGTAATAGAGGACATGAC 5465
972 eSerLeuGluLysGluLysPheLysAlaIleThrgLysGluLeuLys 989
5466 ACCCTGTGTGTGACCAAGAAATATGAGGCGCACGAAAGCGATCAG 5515
989 ySGluLysLeuSerGlnIleThrgLysIleThrgSerGlu... 1003
5516 ACCCAACAAATAGAGAAAGCATGACTGAGAAATACCTTAAAGAGCT 5565
1004 ...GluGlu 1005
5566 AGAGCCGCTGAGAAAGCTGATGAAAAGAAAGACAGCTGTGCTTAAACA 5615
1005 uAlaLysLysCysLysGluGluAsnAspLysLeuLysLysGluIleLeu 1022
5616 ACTGAAAGAAAGTGAAGCATGACATGATTTTACTTAAAGGATTAAGTGGAGA 5665
1022 hLeuGluLysAspLeu...LysAspLysAsn 1031
5666 ACCTTGAAGAGAGCTGAGATTAACCCAGGAAACAAACACATGACACT 5715
1032 ValHisIleGluLysAsnSerTyroIleThrgLysArgAlaLeuSerArgLys 1048
5716 CTGAGGCAAGAAATATGCTCAAGAGAGAGCTGAGACCTTAAAGAGCAAGAA 5765
1048 rGluGluLeuAsnArgIleLysAspLysLeuGluLysIleThrgLys 1065
5766 AGAAGGATGACCAAAAGCTCTGAGAGCTGCAAGATTAATGATGTTGACT 5815
1065 LaLysLysGluLysGluLysLeuValGluLysAsnAlaLysGlnHisSer 1081
5816 TAAAGTCAAGAAAGAAAGATCTGCAAAATATATCAAAAGAGCAAGAG 5865
1082 GluIle...LeuAlaIleGlnThrLeuLeuGlu 1091

```

```

6679  C G A G T T G C A T C C A A A A A A A C A G C T G A A A A A A A C A G G A A A C T G A G A 67229
1344  n g i n t a l t a r s e r l e u g i n c i n l e u l a s p r l i a s p a r t g c i n h t s 1361
6729  G G A A T T A C C T G A M A T T A C T A C T A C G C G T T C T A M G C T G A A A A A A A C C 6778
1361  i n c l u a l . . . . . i l e a l i l e 1366
6779  A G C C T C T C T T T T G A C A C A A C A A C A G T A A G A G T A A A A T C T C A G A 6828
1367  T T A T A T T T H T L e u l e u s e r l a l a c i l o g l i y n s e c a r p c i l s p a y a 1388
6829  T A C C A G A G A A T T G A C T T A A A G A A G A T G T C A C T C A C A G A C T 6878
1383  I g i n l a l i l e u l e u g i n 1389
6879  G G A G A T A C C T T T T T A A A 6897
seq_name : /cgn2_6/ptodata/2/lin/5A.COMB.seq; US-08-139-937-12

```

```

seq.document.location:block:
    Sequence 12, Application US/08139937
    Patent No. 5621070
GENERAL INFORMATION:
    APPLICANT: ABBOTT LABORATORIES-IRMA
    APPLICANT NAME: SHAN, BETI
    TITLE OF INVENTION: CELLULAR GENES ENCODING RETINOBLASTOMA-ASSOCIATED PROTEIN
    NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
    ADDRESSEE: CAMPBELL AND FLORES
    STREET: 4370 LA JOLLA VILLAGE DRIVE
    CITY: SAN DIEGO
    STATE: CALIFORNIA
    COUNTRY: USA
SEQUENCE INFORMATION:
    CONTAINER SPINDABLE FORM:
    MEDIUM TYPE: floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/139,937
    FILING DATE: 20-OCT-1993
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
    APPLICATION NUMBER: US 07/979,156
    PRIORITY DATE: 20-OCT-1992
ATTORNEY/AGENT INFORMATION:
    NAME: CAMPBELL, CATHERYN
    REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9370
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 619-535-9001
    TELEFAX: 619-535-8949
SEQUENCE CHARACTERISTICS:
    INFORMATION FOR SEQ ID NO: 12:
        LENGTH: 155
        STRAND: single
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
        MOLECULE TYPE: cDNA
US-08-139-937-12
Alignment_scores:
    Quality: 528.50      Length: 1256
    Rello: 67/350       Gaps: 155
Percent Similarity: 55.301   Percent Identity: 21.46
alignment_block:
US-09-750-599-2 x US-08-139-937-12 ..

```

```

Align seg 1/1 to: US-08-139-937-12 from: 1 to: 488

276 Aaenclnargclunhlsrlnaslnlecln.....Aerleuagclunlclglns 290
|||||.....:|||||:|||||
39 AATCGAGAAAGATATATCAATATTCACCTCCGGGTAAAGAGACATCAAA 88
|||||.....:|||||:|||||
290 nclunAsrleuLyS.....GluatgleuLySllle 301
|||||.....:|||||:|||||
301 InclInclunlInglleleuLeuAsrlyuAsnIngluenglnleuIn 317
|||||.....:|||||:|||||
139 AAGGT.....TTGGTAAATGAAATTAAGAAATATGTGACATCAAA 176
|||||.....:|||||:|||||
318 LeuAsn...GluGluValMetValAlaAsrAsrleuclnserclunlsgcl 333
|||||.....:|||||:|||||
177 CTCTATTTAAGAGAGATACATTAATGACAAATTAAGAAATGACATCA 226
|||||.....:|||||:|||||
333 uLyseuLySserleuLeuAlaLalysgluysgnlnleuInclunsercl 350
|||||.....:|||||:|||||
227 A...TTGGAATAATAGTTGGGCACTTAAGAAAGAAATCAAGAT.... 269
|||||.....:|||||:|||||
350 euArgrhrllleclunlaleuLySserAsrPheLyStryPheGluSerAsp 366
|||||.....:|||||:|||||
270 .....TTAAGTAAATATGAAATATTTGTCTGTAT 302
|||||.....:|||||:|||||
367 HlsleuclnlySerGlyserHlsPheArglySglunSmetleuclnlysercl 383
|||||.....:|||||:|||||
303 CAC.....CAGAGATTACTCCAGAG... 323
|||||.....:|||||:|||||
383 nclunlmetlyrmetThrAsrSerGlnCythrSerThrclunlmetProy 400
|||||.....:|||||:|||||
324 .....GTGAAGACTTGTGAAGGCTCAATCTGTATTTAGAA 360
|||||.....:|||||:|||||
400 alHlsMetGlnSerAsrSerMetleuArgProleuclnleuAlaLeuPro 416
|||||.....:|||||:|||||
361 TGCATGACAGATTAATCATCATCATGACATATATGAGATATATGGCCAG 410
|||||.....:|||||:|||||
417 AaenclnAlaSertrySerGlnaenclunle.....leuLySglunleucl 432
|||||.....:|||||:|||||
411 GTCAATGACAGCTGACAGAGAGATATTTCTTGATGGAATTAAGACCTGAG 460
|||||.....:|||||:|||||
432 uAlaMetAsrgrhrPheCyAsrSerAlaLySglInAsrArglyleucln 449
|||||.....:|||||:|||||
461 TACGATCAGATCGGAGAAAGCTTACGATTGACATGAGACCTCTACTGCG 510
|||||.....:|||||:|||||
449 lnaenclunleuAlaHlsLySValAlaGluCyLySValAlaLeuAlaGlu 465
|||||.....:|||||:|||||
511 AGGCTGACTTA.....CAGGTACGTCAACAGAGACATATGTTT... 551
|||||.....:|||||:|||||
466 CysglunllyValleuclnAsrSerAsrlyuclnlnleuLySglunleucln 482
|||||.....:|||||:|||||
552 ...GAAAGAGACATCAAAATTAAGCAGAGCTTATTCCTGCTTGAAGA 596
|||||.....:|||||:|||||
482 pAlaLeuLySAsrValGlnLyAsrMettryclunserclunlsglyVal 499
|||||.....:|||||:|||||
599 AGCACTCTGACGTCTCAGAGTGAAGAAACCACTCTGTGGAGAAATGG 648
|||||.....:|||||:|||||
499 ArglnMetGlnlnrHlsPheleuAlaLeuLySglunHlsleuThrSerAsp 515
|||||.....:|||||:|||||
649 ATACTATG.....TCAAAA 662
|||||.....:|||||:|||||
516 AlAlaLathrGlyAsnHlsAsrLeuMetGluGluLeuLySAsrGlnleu 532
|||||.....:|||||:|||||
663 AAACACAGCGACATGATCAGTTTCGTGAAAAATAAGAGAGAAACAAACA 712
|||||.....:|||||:|||||
532 AsrPheLySValLyStryclunlsglyAlaSerAlaGluValGlyLySleu 549
|||||.....:|||||:|||||
713 AGAGCTTGAGCTTCATCAAGTGAAGTCTGCTCAATCCATCTCAAGCTG 762
|||||.....:|||||:|||||
549 rGAsnGlnlnleuLySglnAsnGlnMetleuValGlnleuPheLySAsr 565
|||||.....:|||||:|||||
763 AGCGAGAGGTGAGAGAAAGAGACGAACTCTTGAGACTTTGTCTGTGAT 812
|||||.....:|||||:|||||

```

```

566 GluGlyLySleuMetclunlunlnlyAsrGluclnlySglunleuSerne 582
|||||.....:|||||:|||||
813 GTGATGACCTTTTAAAGACACAAAGATCTCAATCTCAAGATGAGAG... 860
|||||.....:|||||:|||||
582 lCySglunleuclnArgGlyAsrGlyValGlyLySleuThr.....GluM 597
|||||.....:|||||:|||||
861 ...AGTTGCAAAAGAGACTTCACAGGCACTCTTTTGCAAAATGTGACC 906
|||||.....:|||||:|||||
597 etclunllyclnleuLySAsrleuSerAlaLySleuAlaLeuSerllePro 613
|||||.....:|||||:|||||
907 TGAAGAACCAATTTGACACATGAAATTAAGAAAGAAATGTGCTGTCTG... 953
|||||.....:|||||:|||||
614 AlaGlnlyPheclunlunSmetlySerleuSerleuclnleuclnleucl 630
|||||.....:|||||:|||||
954 ...AAGCAATCTCAAAAGCTGACAGGCAAGACTGATGATGATCAATATGA 1000
|||||.....:|||||:|||||
630 uLySAlaLySleuLleAsrPValGlnArgGlyUtryclunlAsrSerleu 646
|||||.....:|||||:|||||
1001 AAGG.....CTGAATGTCTCAAGGCTTGAGAGAGCCCATGCG 1038
|||||.....:|||||:|||||
647 .....AaenclunlThr 649
|||||.....:|||||:|||||
1039 TCGAAGAAAGGTGATTCGCTTGAAGGCTCATCAACACAGAGAAAGTG 1088
|||||.....:|||||:|||||
650 ArgProleuLySArgGlyleuclunlunlnleuLySAlaLySleuAlaGlnl 666
|||||.....:|||||:|||||
1089 CATACGCTGAGAGAGAGCATCGAAGAAATCGAAGATTCGCTTGAAGGCGCA 1138
|||||.....:|||||:|||||
666 sValLySProclunlunl.....GluGlnleuLySAsrArglyleucln 681
|||||.....:|||||:|||||
1139 TGAAGAGAGCACTGCACATCGAGAGAGAACTGAAAGAAAGCC...GAGC 1185
|||||.....:|||||:|||||
681 lInLySserclunlunleuclnlySArglyleuclunlunleuThrlySerlyAsn 697
|||||.....:|||||:|||||
1186 GAGAAATATCTCACTTAAGCTTAAGATGAGAAAGCTTGAAGAGAAATGG 1235
|||||.....:|||||:|||||
698 GlnThrleuGlnlySglunlclunllySValLySglunAsrPAsnlyleucl 714
|||||.....:|||||:|||||
1236 CAGATCTCAAGAAACCAACAGAGCTATGATGTTCTTGATCCGAGATTC 1285
|||||.....:|||||:|||||
714 uThrGlnGlnValAsnAsnleuclnThrclunlmetlySAsnValProleucl 721
|||||.....:|||||:|||||
1286 CAAGCGAGATGAGAGACTTAAACCAACAAATTA..... 1319
|||||.....:|||||:|||||
721 ySValSerclunlclunlmetlySserHlsAsrVallleuAsrAsp... 746
|||||.....:|||||:|||||
1320 .....GAGAGATGGCCAGAAAGCTTGAAGATTTTGAATTAATGACT 1361
|||||.....:|||||:|||||
747 .....leuAsnlyLySleuSerAsrVallyHlsLyStryThrclunl 761
|||||.....:|||||:|||||
1362 GTCACTTAAGGTCTGAAAAAATAATCGCAAAACAAATCAAGAAAA 1411
|||||.....:|||||:|||||
761 styS.....leuclunetclunllySleuclunleuclunleuclnleucl 775
|||||.....:|||||:|||||
1412 ACAAGCTCAAGTTCTCAATACAGAGAGTCTCTCTCTTAAAGATG 1461
|||||.....:|||||:|||||
775 euSerlySAsnValSerArgleuclunlThrValHlsleuclnleuclunl 791
|||||.....:|||||:|||||
1462 TGTTAGAGAAAGAGACCAAGAGAG...ATACAGATCAAAAGAAAGAACT 1508
|||||.....:|||||:|||||
792 HlsGlyLySglunetMetAlaLeuLySserAsnllleThrclunleuLySly 808
|||||.....:|||||:|||||
1509 .....AaenclunlThr 811
|||||.....:|||||:|||||
808 sGlnleuSerclunleuclunllyS.....CysGlyLyAsp 821
|||||.....:|||||:|||||
1535 TCACTTAAGAGACTTAATGAGGAGCTGAGAGCTTGTGTGTGATCCAG 1584
|||||.....:|||||:|||||
821 lIn..... 821
|||||.....:|||||:|||||
1585 AATTTATGAAGGCACAGACAGAGCTTAAAGCCACCAATTAAGAGAGAG 1634
|||||.....:|||||:|||||

```

```

822 .....GluThrIleTyrSerIleuMetSerGln 831
1635 CATCAAGCTGAGAAATGACATTGAAAAGCTGAGAGAGCCCGCTAGAAAGCTCA 1684
831 naaNaapLeuIleuTyrIleMetSerIleHsGlnIleProvalIleTyrIleHn 848
.....
1685 TGAAAAGAGAGAGCTGTCTCTCTACAAAGAGAGCTGAGAAAGAGAGATC 1734
848 IAGGAGIleuIleuTyrIleAlaIleuSerSerThrLeuAsp..... 860
1735 ATCCAGATTACTTAAAGGTGAGAGGTGAGAGAGAGCTGTCAAGAGAGCTAGAG 1784
861 .....LysThrAsnArg..... 864
1785 ATAGCCAGAGCAAAACCAAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1834
865 .....GluLeuValAspValIleuIleuTyrGluAspIleleuGln... 878
1835 AGGAGAGGTGAGAGAGCTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1884
879 .....GluThrValIleuIleuTyrAspGlnIleuIle 889
1885 TGAGAGCTGTGAGATTAGATGTGTACTATAGAGCTGAGAAAGAGAGAGAG 1934
890 LeuIleuTyrAsnIleuGlnIleuAspThrIleuAsnIleuVal...LysAlaGluTyr 905
1935 CTGACAAATGAAATTAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1984
905 rIleoser.....LeuATGlnIleuGlnIleuIleuMetS 916
1985 AATTAATTCATCATTTGAAAATTTTTCAGAGAGAGAGAGAGAGAGAGAGAG 2030
916 eGluIleuArgIleuSerMetIleuIleuValGlnIleuAspSerAlaIleuIle 932
2031 .....AAAGTACAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2072
933 Leu...AlaIleuTyrIleuIleuSerGlnIleuGlnIleuValIleuIleuHnS 947
2073 CTTCAAAG 2122
948 .....GluGlnIleuAlaIleuGlnIleuAspGlnIleuAspThrIleuGln 962
2123 TGACCAAG 2172
962 LysGlnIleuIleuTyrIleuIleuTyrAlaProIleuIle 972
2173 AGTGTCTTGAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2222
973 .....Se 973
2223 AAAAATTAATTAATTTGTTTCATCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2272
973 rIleuGlnIleuGlnIleuGlnIleuIleuIleuIleuIleuIleuIleuIleu 987
2273 AGTAAAGATGAG 2322
988 .....LeuIleuGlnIleuIleuSerGlnIleuIleuIleuIleuIleuIleu 1000
2323 GAGTGAAGAAATCAATTCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2372
1001 ThrSerGlnIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleu 1017
2373 CAGGTGAG 2422
1017 eLysGlnIleuIleuThrIleuGlnIleuIleuAsp.....LeuIleuAspIleu 1031
2423 AAATCTGACAGGTGAGATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2472
1031 gnaValIleuIleuIleuAsnSerTyrGluThrIleuAlaIleuIleuIleuIleu 1047
2473 ATGCTGTCTGTGAG 2522
1048 ThrGln.....GluLeuAsnArg..... 1053

```

```

2523 CTAGAGATGAGCTGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2572
1054 .....GlnIleuIleuAspIleuGlnIleuTyrTyrThrGlnIleuIleuIleuIleu 1070
2573 AAAAGTAAACAAATGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
1070 IuleuIleuValGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1086
2623 ATGAGATGAG 2672
1087 GlnIleuIleuGlnIleuGlnIleu...HisValProIleuGlnIleuValGlnIleu 1102
2673 AAAAATGAGCTGAG 2122
1102 rIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1116
2723 CAGCAAAATGATCATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2166
1118 eLysIleuIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1129
2767 TGAG 2816
1130 .....GlnIleuValThrIleuIleuIleuIleuIleuIleuIleuIleuIleu 1143
2817 AAAGTACAG 2866
1143 sAsnSerSerValProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1160
2867 AAAG 2910
1160 IuleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1176
2911 AGGTGAG 2960
1177 SerGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1193
2961 CTGAGTTCAG 3010
1193 rThrIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1210
3011 GCTCAATTAATTCATGAA.....GCTAGTACTGAGAT.....TTG 3048
1210 eLysIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1226
3049 AGGATTTGAG 3074
1227 AAGGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1242
1243 AATTCAG 3124
1245 .....GlnValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1257
3125 GAAATTAAGCTTTGATCAATTCCTGTAAACAGCTG.....GAGAGAG 3168
1257 IuleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1273
3169 AAAAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3218
1274 Glu.....ArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1281
3219 AGCAG 3268
1281 rThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1298
3269 TGAGATCAAG 3314
1298 IuleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1314
3319 ATGATTAATCTGAG 3368
1315 LysGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1330

```

```

3369 GAGAAAGCTAAAGAGATGTTAGACACACAGTGGCCATCTCTGTTCACA 3418
1330 rProserLySarGlnSerGln..... 1337
3419 GCACCTTAACACAGATTCGCCAGGGCTCTCTTCTCTAGGTCACAGTCTTC 3468
1338 .....Leu11easrSerLeuGlnGlnValaTsrGln 1350
3469 CAGACCATCTCCAAATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 3518
1351 GlnGlnLeuAlaAspAlaAspArgGlnHisGlnGlnValaIle 1364
3519 CAAATTAAGCTTCAGCCAGACGACAAAGATTCACAGCAATA 3560

seq_name: /c9n2.6/ProdData/2/lna/PCTUS.COMB.seq:PCT-US93-11310-12
seq_documentation_block:
? Sequence 12, Application PC/US9311310
? GENERAL INFORMATION:
? APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
? TITLE OF INVENTION: CELLULAR GENES ENCODING
? NUMBER OF REFERENCES: 14
? CORRESPONDING ADDRESSES:
? ADDRESSER: CAMPBELL AND FLORES
? STREET: 4370 LA JOLLA VILLAGE DRIVE
? CITY: SAN DIEGO
? STATE: CALIFORNIA
? COUNTRY: USA
? ZIP: 92122
? COMPUTER READABLE FORM:
? INVENTOR: TERRY, DORRIS L
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/11310
? FILING DATE: 19-NOV-1993
? CLASSIFICATION:
? ATOMASER: INFORMATION:
? NAME: CAMPBELL AND FLORES
? REGISTRATION NUMBER: 31 815
? TELECOMMUNICATION INFORMATION: FP-CI 9790
? TELEPHONE: 619-535-9001
? TELEFAX: 619-535-8949
? INFORMATION FOR SEQ ID NO: 12:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1264
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? PCT-US93-11310-12

alignment_scores:
? Overall: 528.50 Length: 1264
? Ratio: 0.756 Gaps: 53
Percent Similarity: 55.301 Percent Identity: 21.440

alignment_block:
US-09-750-590-2 x PCT-US93-11310-12
Align seq 1/1 to: PCT-US93-11310-12 from: 1 to: 4868
276 AaenGlnArgGlnHisGlnGlnGln.....AspLeuGlnIleGlnAs 290
|||||.....|
39 AATCAGCAAGATTCATCAATCTTCAACTCGGGGTAAGAAAGACATCAAA 88
|||||.....|
290 nGluAspLeuIlyS.....GlnArgLeuArgIlySLeuG 301
|||||.....|
89 TGCAATTTAGATTAAGTACTCATGATGATAGAGACCGCTGACAGAAAGTTG 138
|||||.....|

```

```

301 LngGlnGlnArgIleLeuLeuAspIlyValaAsnGlyLeuGlnLeuGln 317
|||||.....|
139 AdGgt.....TTCGTAATGAAGAAAGATTAAGCTCAAAA 176
|||||.....|
138 LeuAsn.....GlnGlnValaMetValaAlaAspArgLeuIlySGL 333
|||||.....|
177 CTCATTTCAGAGAGTACACACTATGACCAAAATTCAGACATCATAGA 226
|||||.....|
333 IlySLeuIlySerLeuLeuAlaAlaIlySGLIlySGLIlySGLIlySGL 350
|||||.....|
227 A.....TTGGAAAAATTAAGTGGCAACTTAAGAAAGAAACATCAAT... 269
|||||.....|
350 euArgThrIleGlnIleuIlySerArgPheIlyGlyIlySLeuIlySAsp 366
|||||.....|
270 .....TTAAGTCAAAAATTTGAAATATTTTCTCTGAT 302
|||||.....|
367 HisLeuGlnSerGlySerHisPheArgIlyGlnIlySLeuLeuIlySGL 383
|||||.....|
303 CAC.....CAGAGTTCCTCCAGAG... 323
|||||.....|
383 nGlyGlnMetIlyMetPheAspSerGlnGlyIlySerThrIlyMetPhe 400
|||||.....|
324 .....GTACAAACTCTGAAGCCCTCAATTCATTAATTA 360
|||||.....|
400 aHisMetGlnSerArgSerMetLeuArgPheGlnGlnLeuAlaLeuPro 416
|||||.....|
361 TCATTCACAAATTAATCATCATCGAATTAAGTATTAATTCGCCAAG 410
|||||.....|
417 AaenGlnArgGlnSerGlnGlnGln.....LeuIlySGLIleuGln 432
|||||.....|
411 GTGATGACAGCTGGAGAGAGATTCCTTGAATGTGGAATACGTGA 460
|||||.....|
432 uAlaMetArgThrPheCysAspSerAlaIlySGLAspArgLeuIlySLeu 449
|||||.....|
461 TAGATTCAGATTCAGAGAGAGATTCCTTGAATGTGGAATACGTGA 510
|||||.....|
449 IlySLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 465
|||||.....|
511 AGGCTGACTTA.....GAGTACTTCAACGAGAACGCTATGTTA... 551
|||||.....|
466 CysGlnArgValIlySGLAspSerArgPheGlnIlySGLIlySGLIlyS 482
|||||.....|
552 ...GAAAGACAAATTAATTAAGCAGAGCTTATTCCTCTGGAAG 598
|||||.....|
482 PAlaLeuIlySAspIlyGlnIlySArgMetIlyGlnIlySLeuValI 499
|||||.....|
599 AGAATCTCAGTGTGACAAATGAGAAACACCTTCCTGCAATTAAG 648
|||||.....|
499 ySGLMetGlnHisPheLeuAlaLeuIlySGLIlySLeuIlySAsp 515
|||||.....|
649 ATACTAT.....TCAAAA 662
|||||.....|
516 AlaAlaThrIlySAsnHisArgLeuMetGlnGlnIlySAspGlnIlyS 532
|||||.....|
663 AAACCGGCGACGATCGATCGATCTCTCAAGCAATTAAGCAACAAAC 712
|||||.....|
532 SasPheIlyValIlySArgIlyValaSerAlaGlnValaIlySLeu 549
|||||.....|
713 AGAGCTTGATCTCATCAAAAGTACGTGTCCATCATCATCATGAGTGCAG 762
|||||.....|
549 rGAsnGlnIlySGLAsnGlnMetLeuValaGlnIlySLeuIlySAsp 565
|||||.....|
763 AGCAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 812
|||||.....|
566 GlnGlyIlySLeuMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 582
|||||.....|
813 GTAGTGTAGCTTTAAAGAGAAACAAATCATCTCCACAGAAAGCTCAG 860
|||||.....|
582 IlySGLIlySLeuGlnArgIlySArgIlySLeuIlySLeuIlySLeu 597
|||||.....|
861 ...AGTTTGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
|||||.....|

```

[illegible]

```

2623 ATGAGATGGCAGAGAAAACACAGAGCTGCAGAGAAAGCACTGACAGAG 2672
1087 GlnThrLeuGlnIuysGln...HisValProLeuGlnIuysGln 1102
1102 AAAGATACGCTAGCTGAGAGCTGCAGCTACTGCTGGAAGAAATTAAGAG 2722
1102 LeuIuysLeuSerLeuSerGly...ThrIleGlnThrLeuIuysGlnIuT 1116
1102 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2723 CAGCAAAAGATCAATTGAGAGAGCTCACACATGAA...AAATGATGAT 2768
1118 euIysThrIuysGlnAGCysTyrGluIuysGlnIu 1129
1118 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2767 TGAGAGAGAGACCTGAGATGCTGACCAAAAGACCTGAGAAAGAGAGG 2816
1130 .....GlnThrValThrGlnLeuArgIuysGlnLeuGlnIuysGln 1143
2817 AAGTCAGAGAGAGAAATGCTGAAATATGACAGCTGACCTGACAGCTGA 2866
1143 sAnSerSerValProLeuAlaGlnIuysLeuIuysGlnIuysGln 1160
2867 AAAGAAACACAGAGCTTTGCT...TTGACACAAAGAAACAGATG 2910
1160 IuysGlnIuysGlnIuysLeuIuysLeuSerLeuArgIuysGlnIuysGln 1176
2911 AGTGAAGATTCAGACATACACAGAGAGAGATGCTGACAGAGATG 2960
1177 SerGlnAsnIuysThrIuysGlnIuysLeuIuysGlnIuysGln 1193
2961 CTCAGCTCAGAGAGAGCTGAGATGACCTTTAAAGCTGTGAAGAGA 3010
1193 rThrIuysGlnAlaLeuIuysIuysGlnIuysThrArgGlnValIuys 1210
3011 GCTCAATATTCATGAA...GCTCACTCTCAAGAT...TTGG 3048
1210 euIysTyrIuysAlaThrIuysSerAspLeuGlnIuysIleSerAsp 1226
3049 AGCAATTCAGAGAAACACAGATGAC... 3074
1227 AsnGluIuysLeuAlaAsnIuysAsnArgIuysGlnIuysGln... 1242
3075 AATCTAAATATGCTG... 1257
1243 .....GluValLeuHisAlaIuysIuysGlnIuysSerAlaIuysAsp 1257
3125 GAAATGAGCTGTGATCAAAATCTCTGTAAGAGCTG...GAGAG 3168
1257 IuysGlnIuysLeuHisAspSerIleGlnIuysGlnIuysAspGln 1273
3169 AAGAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3218
1274 Gln...ArgCysAspIuysSerLeu 1281
3219 AAGCAGAAAGAGTACTGTATGATACCAAGCTGCATATTAATCAAC 3268
1281 rThrIleThrIuysGlnIuysArgIuysIleGlnIuysSerAlaIuysGln 1298
3269 TGCAATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3318
1298 IuysAlaIuysAspAsnIuysIleThrGlnIuysLeuAsnAspValGln 1314
3319 ATGAATCTGATAGTACTGCTCTGCTATTAAGCCTAAAGAGAT... 1337
1315 IuysGlnAlaLeuAsnGlnIuysLeu...SerGlnLeuThrIuysSerGln 1330
3369 GAG 3418
1330 rProSerIuysArgIuysGln... 1337
3419 GCATCTTAAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3468
1338 .....LeuIleAspSerLeuGlnIuysGlnIuysAlaIuysSerLeuGln 1350
3469 CAGAGACATCTCAATCTCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAG 3518

```

---

```

1351 GlnGlnIuysAlaAspAlaAspArgIuysGlnIuysGlnValIle 1364
1351 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3519 CAAATTAAGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3560
seq_name: /cgn2_6/prodata/2/lna/5a_COMB.seq:us-08-466-390-3
seq_documentation_block:
/ Sequence 3, Application US/08466390
/ Patent No. 5686362
/ GENPAT INFORMATION:
/ APPLICANT: UNIVERSITY, CARP
/ APPLICANT: LIPKARD, ORHAM P
/ TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
/ CORRESPONDENCE ADDRESSES:
/ NUMBER OF SEQUENCES: 6
/ ADDRESSES: TESTA, HURWITZ & THIBEAULT
/ STREET: 125 HIGH STREET
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ APPLICATION NUMBER: 08/466,390
/ FILING DATE: 06-JUN-1995
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: PITCHER ESQ, EDMUND R
/ REGISTRATION NUMBER: 27,829
/ REFERENCE/DOCKET NUMBER: MTP-013
/ TELEPHONE: (617) 248-2100
/ TELEFAX: (617) 248-2100
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 6306 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE KEY:
/ FEATURE TYPE: CDS
/ NAME/KEY: CDS
/ LOCATION: 1..6306
/ PUBLICATION INFORMATION:
/ AUTHORS: COMPTON, DUANE A
/ AUTHORS: SELLAK, ILVA
/ AUTHORS: CLEVELAND, DON W
/ TITLE: PRIMARY STRUCTURE OF NIMA, AN INTRANUCLEAR
/ TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
/ TITLE: SEQUESTRATION OF PROTEINS AT MITOSIS
/ JOURNAL: J Cell Biol.
/ VOLUME: 116
/ PAGES: 1395-1408
/ DATE: 1992
/ US-08-466-390-3

```

Alignment scores:

| Quality:            | 514.00 | Length:           | 1541   |
|---------------------|--------|-------------------|--------|
| Ratio:              | 0.662  | Gaps:             | 61     |
| Percent Similarity: | 50.357 | Percent Identity: | 20.052 |

Alignment block:

US-09-750-590-2 x US-08-466-390-3

Align seg 1/1 to: US-08-466-390-3 from: 1 to: 6306

```

185 lvsghlnasnafrtgh.r.laleumellucylucylurtyrlycyulvayns 201
    .....
643 csmkrtgkmgkccstgcmamkccsttgcctgcmamkmgkmtatgkgh 692
    .....
201 palvalglulval..leulvaynsnclvalalavaylthrlthleulva 217
    .....
693 tgcagctgcagctgcagctgcagmaccgacccctgcagcagmaccg 742
    .....
217 spalaleuldlvhlslasrserferty.rly.r.lalagrlegllyarapnleu 233
    .....
743 atgcga.....csmkrtgkmgkccsttgcctgcmamkmgkmtatgkgh 774
    .....
234 aprrtleuethrlthleulvaylthrlthleulvaynsr..... 246
    .....
775 gacccgctgcagctgcagctgcagmaccgacccctgcagcagmacc 824
    .....
247 .asnlvsglyargluclutfrlyulvsglyrproserleuglnulnlyr 262
    .....
825 csmagacacttgcagacccctgcctgcmamkmgkmtatgkghccgc 874
    .....
263 .....aprrtleuethrlthleulvaynsrlyulvaynsrlyulvaynsr 276
    .....
875 tgcctgcmamccctgcagcagctgcagcagc...ctgcamaccgacmaccg 921
    .....
277 glulagrqlunhlsgrlnamnlleglnarpleucluleglnuglnuaple 293
    .....
922 cagatgcgat...csmamamamamamamamamamamamamamamamam 968
    .....
293 ulvsglulagrqlunhlsgrlnuglnuglnulnagrleleleulvaynsr 310
    .....
969 ttccttttmaactgcgagagtttgcagctgcagcagcagctgcagcagatg 1018
    .....
310 yavvalasnglyleu..... 314
    .....
1019 cccctcamaatgcagctgcagcagcagcagcagcagcagcagcagc 1068
    .....
315 .....gluluglnleulnaglnuglnulvaylmetvalalavaynsrleugl 329
    .....
1069 gacagacgacccgacgacgacgacgacgacgacgacgacgacgacg 1112
    .....
329 usertglulvay...glulvayleulvayserleulvaylvalalvaynsr 344
    .....
1113 ggcacagaaagacccctgcagmaccgacmaccgacmaccgacmaccg 1162
    .....
344 yaglnhlsgrlulvaynsrleulvaylthrlthleulvaynsr 354
    .....
1163 cagacgctgcagmaccgacttgcagctgcagcagcagcagcagcag 1212
    .....
354 ..... 354
    .....
1213 maccgagacgctgcagctgcagctgcagcagcagcagcagcagcag 1262
    .....
355 .....alaleulvaynsrlyarpleulvaynsr 362
    .....
1263 ggcagcagcagcttgcctgcamamamamamamamamamamamamam 1312
    .....
362 yrrpleulvaynsrlyarpleulvaynsr.....leuglysergl 371
    .....
1313 tgcctgcagmactgcagcagcagcagcagcagcagcagcagcagcagc 1362
    .....
372 serghrplearglvaygluvarpleleulvaynsrln..... 383
    .....
1363 ggcagcagctgcagmaccgacmaccgacmaccgacmaccgacmaccg 1412
    .....
383 ..... 383
    .....
1413 csmagacactgcagctgcagcagcagcagcagcagcagcagcagc 1462
    .....
384 .....glulvaynsrlyarpleulvaynsr 390
    .....
1463 cccctgcagcagcagcagcagcagcagcagcagcagcagcagcagc 1512
    .....
391 serglncylvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 407
    .....

```

```

1513 tgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1562
    .....
407 leulvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 424
    .....
1563 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1612
    .....
424 serglncylvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 440
    .....
1613 cc.....ctgcamaccgacmaccgacmaccgacmaccgacmaccg 1656
    .....
441 alalvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 457
    .....
1657 tgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1694
    .....
457 aslucylvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 474
    .....
1694 cagatgcagmactgcagcagcagcagcagcagcagcagcagcagcagc 1744
    .....
474 serglncylvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 479
    .....
1745 cccgacmactgcagcagcagcagcagcagcagcagcagcagcagc 1794
    .....
480 leulvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 496
    .....
1795 cccgacmactgcagcagcagcagcagcagcagcagcagcagcagc 1844
    .....
496 ulvaynsrlyarpleulvaynsrleulvaynsrleulvaynsrleulvaynsr 513
    .....
1845 csmagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1888
    .....
513 hrserarplalathrlthleulvaynsr..... 520
    .....
1889 cccgacmactgcagcagcagcagcagcagcagcagcagcagcagc 1938
    .....
521 .....hlsagrleulvaynsrlyarpleulvaynsrleulvaynsr 535
    .....
1939 gacgacgacccgacmaccgacmaccgacmaccgacmaccgacmaccg 1988
    .....
535 yvalvaynsrlyarpleulvaynsrleulvaynsrleulvaynsrleulvaynsr 552
    .....
1989 csmagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2038
    .....
552 leulvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 565
    .....
2039 tgcgctgcagcagcagcagcagcagcagcagcagcagcagcagc 2088
    .....
566 glulvaynsrleulvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 582
    .....
2089 maccgacgacgacgacgacgacgacgacgacgacgacgacgacgac 2138
    .....
582 tcyagclvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 597
    .....
2139 csmagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2185
    .....
597 serglncylvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 613
    .....
2186 tgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2220
    .....
614 alaleulvaynsrlyarpleulvaynsrleulvaynsrleulvaynsr 630
    .....
2221 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2252
    .....
630 ulvaynsrlyarpleulvaynsrlyarpleulvaynsrlyarpleulvaynsr 647
    .....
2253 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2302
    .....
647 sn.....glulvaynsrlyarpleulvaynsr 651
    .....
2303 agcctgcagcagcagcagcagcagcagcagcagcagcagcagcagc 2352
    .....
652 leulvaynsrlyarpleulvaynsrlyarpleulvaynsrleulvaynsr 668
    .....

```



```

2353 CTCGCGCGGAGCTG.....CGAGGCGCATCGCTCCGACGACGACG 2396
668  APTGGLUHLNLSGLULULeuLysSerArgLeu..... 679
2397  TCAAGTATAGTGTAGAGACTCTCCAAAGAAATAGCTGCGGCTGACG 2446
680  .....GluGlyLysSerGlyLysGluGlyLysArgGlyLeuThr 691
2447  GGTATGAGCATGCCAGACAGAGAGCGACATATGCGCATCTCCAG 2496
692  GluLeuThrSerLysAsnGlnThrLeuGlnLysGlnLeuGlyValGly 708
2497  GAA.....CAGCTGATGACTTTGAAAGAGAGATGTGAGAGAGCGG 2537
708  AlaLeuAspLysLysLeuThrGlnGlnValAsnAsnLeuThrThrGlu 725
2538  CCAAGAGCTCCGAGAGCGCAAGAGAAAGTGGCGAGCTCAATACCA 2587
725  GCTGAAATValProLeuLysValSerGlnLysLeuLysLysSerHis 741
2588  GCGAG.....CTCCAGATAGCCGCGACCGACGAAACAA..... 2619
742  ValLeuValAspAspLeuAsnLysLysLysSerAspValThrHisLys 758
2620  .....CTAGTGAAGTCCATGCGCACTGCGACGAGCGCTCCACAG 2663
758  ThrGlnLysLysLeuGlnLeuSerGlnLysLeu.....L 769
2664  CCAAGAGAGAGAGTACGGCCCGAAGCGCTTGCAGATACCTGCTC 2713
769  euMetGluAsn...AlaSerLysSerLysAsnValSerArgLeuGlnThr 784
2714  TCCAGGAAAAGTGTGCTGCACCGACCAAGAGTGGCCCGCTTGAGAG 2763
785  ValPhe..... 786
2764  TTGCTCCGACAGCGAGTGCACGACGAGAAACAGCTCCGCGAGTTACT 2813
787  .....LeuProGlnArgHisGlnLysG 795
2814  CAAAGAACCTCGAGAGCGACAGACAGACCGCGAGTGTGAGAGAGC 2863
795  Lu.....MetMetAlaLeuLys 800
2864  AACAGAGAGCGCAGTTCTGCACACACAGAGCAGCGCTGACAGTATG 2913
801  SerAsnLeuThrGlnLysLysGlnLeuSerGlnLeu..... 813
2914  CCGGAGCGACAGCAAGTGGCAATGATGACATGACAGCTGCGGCGCGCT 2963
814  .....AsnLysLysCysGlnLysAspGlnGlnL 823
2964  GATGAGAGACCGAGGCGACGACAGAGAGAGCTGCGACAGAAAGG 3013
823  ysiLeyrSerLeuMetSerGluAsnAsn..... 832
3014  AGTGTGCGGCGGCTACCCAGAGAGCGAGGCGCTCCCGAGCTGACCTT 3063
833  .....AspLeuLysLysThrMet 838
3064  CTGAGAGAGCGCGCCAGAGACAGCTTGACATGATGCTGACAAAGCG 3113
838  eSerHisGlnArgValProValLysThrHisGlnGlnLysLysThrAl 855
3114  CAAAGCAAGAGCTGTGAGATGCTGACCTCCAGACAGACAGCTGCTATG 3163
855  euSerSerThrLeuAspLysThrAsnArgLysLeuValAspAlaLys 870
3164  CCGTACAGAAAAGAGAGCGACAGAGCAGAGGTGGCGCAAGCTTCTGT 3213
871  .....LysLysCysGlnAspHisAsnGlnLysGlnVal 881
3214  CTGAGAGCGAGCCAGATTAAGAGAGCTGTGAGAACTTGCGAAACCTGA 3263

```

---

```

881  LysLysLysAspGlnAsnGlnLysLeuLysArgValAsnLeuGlnAsnThr 898
3264  GCAAGTCAAGAA.....CAGCTGCGTAAAGAAAGAGAGAGCGACT 3307
898  LysAsnGln..... 900
3308  CTGCGTACAGAGCCCAATGTAGCGTCTGCGACAGACAGCGAAAGCG 3357
901  .....ValLysAlaGlnThrLysSerLeu..... 908
3358  CCCAAGCTGAGAGCACTCGCGCGAGAGAGAGAGAGCGATCGAAGCATG 3407
909  ArgGlnLysGlnLysLysMetSerGlyLysValLysSerMetLysLys 925
3408  CCGAAGACAGAGAGAGCGAGCTGACAGCTCGAAGCGACGCTCGAGGCT 3457
925  aGlnAspAsnSerAlaGln.....LysLeuAla 934
3458  AGCGAGCGCTCCGAGCTGACCGCGACAGCTGCTGAGACCTCGAGAG 3507
935  LysThrLysLysSerGlnGlnGlnLysLysValThrLeuAsnGlnGlnAl 951
3508  CAGTTAGAGAGAGAGCGCCAGAGAGCTGAGAGAGAGCTGACAGTGC 3557
951  AlaGlnLysArgGlnLeuLysAspThrThrLeuGlnLysGlyLe 964
3558  CTCGAGCCAAAGCGAGTGTGAGTGTGCTTCCGACCAAGATACAAAGCA 3607
964  ..... 964
3608  CCAAGCGTGAAGATAGTGAAGAGCCCAAGTGCCTCCGCGCGCGAAG 3657
965  ...LysLeuLysThrAlaProLeuLeuSerLeuGln..... 976
3658  GCTGAGAGAAAAGATACCTCATACAGCTGAGCTGAGAGAGAGTGCAT 3707
977  CysGlnArgLysPheLysAlaThrGln.....LysGlnLysLysG 990
3708  CCTGATGCGCAGTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGATGAG 3757
990  LysGlnLysSerGlnGlnThrGlnLysThrAsnThrSerGlnGlnGlnAl 1006
3758  GAGCTGTATAGCGCGAGTCAAGAGAGAGAGAGAGAGCTGAGAGAGCTG 3807
1007  LysLysCysLysGlnGlnLysAspLysLysLysLysGlnLysGlnLys 1023
3808  GCTGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3857
1023  Gln.....LysAspLeuLysA 1029
3858  AGCTCTCTCTCCGAGAGAGTGCAGAGCGCTCCGAGAGAGAGAGAG 3907
1029  spLysAspValHisLysGluAsnSerThrGlnThrAlaLeuLeuSer 1045
3908  AACAGCGAGTGTGCTTCAAGAGAACCTCGAGAGAGAGAGAGAGAGCT 3957
1046  ArgLysThrGlnGlnLysLeuAsnArgGlnLysLysAspLysLeuGlnLys 1062
3958  CAGCGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4007
1062  ThrGlnLysLysLysGlnLysGlnLysLysLysLysLysLysLysLys 1078
4008  CTTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4057
1078  ysiGln.....ThrSerGlnLysLeuAlaLysGlnThrLysLeuGln 1091
4058  CAGAGCGCTGTGATGAGTGTGCTGCGACTTGAAGCAAGCTTCCAGAG 4107
1092  .....LysGlnHisValProLeuGlnGln 1099
4108  CTCAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4151

```

```

          |||      |||||            |||||||             |||||              |
977       CGCGACGCTGGCCAGCTGCAGTGTGAAGCAGCAAGGTTGCCATTS    5026
1356     LAAPAAAGTGTAAGTCAGTGTGTAATT   1363
           |||||             |||
5027     CAGCACGACGACTGCAGACCTGG    5049

seq_name = /cgn2_6/podata/z/jna/5A.COMB.seq;us-08-470-950-3

$$$documention.block:
Sequence 3, Application us/08470950
Patent No. 5698433ION:
GENE ORIGIN INFORMATION:
APPLICANT: LIGCARD, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: TESTA, HORMITZ & THIBAUT
STREET: 1155 HIGH STREET
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATOR: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE APPLICATION: Genbase #1.0, Version #1.25
COMPILATION METHOD: GENBASE COMPILER
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESO, EDWARD R
REGISTRATION NUMBER: ZI,829
REFERENCE/DOCKET NUMBER: MPF-013
TELECOMMUNICATIONS INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7000
INFORMATION FOR SEQ ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6306 base pairs
TYPE: nucleic acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
NAME/KEY: CDS
LOCATION: 1..6306
PUBLICATION INFORMATION:
AUTHORS: COMPTON, DUANE A
AUTHORS: SZILAK, ILVA
AUTHORS: CLEVELAND, DON W
TITLE: PRIMARY STRUCTURE AND NUMA AN ITFRANUCLEAR
TITLE: REGULAR EXPRESSION OF NOVEL PATHWAY FOR
JOURNAL: J Cell Biol.
VOLUME: 116
PAGES: 1395-1408
DATE: 1992
US-08-470-950-3

Alignment scores:
Quality: 514.00 Length: 1541
Ratio: 0.662 Gaps: 61
Percent Similarity: 50.357 Percent Identity: 20.052

```

Align seg 1/1 to: US-08-470-950-3 from: 1 to: 6306

```

185 LysGlnAsnArgThrAlaMetLeuLysGlyGlyLeuLysAsp 201
    ::::::::::::::::::::
643 CAGATGACGCGCGTGAAGACGACCTGCTGCGAGGAAATTAATAGCGA 692
    ::::::::::::::::::::
201 PAlaValGluLys...LeuIleLysAnglyAlaAspValThrLeuLys 217
    ::::::::::::::::::::
693 TGAGCTGACGCTGAGCTAGCTGAGAACGCCAACCTGACGAGAAAG 742
    ::::::::::::::::::::
217 sPaIaLeuGlyHisAspSerSerTyrTyrAlaArgTyrIleYAspAsn 233
    ::::::::::::::::::::
743 ATCCA...CAGATGACATGATGACACGCGCATTT 774
    ::::::::::::::::::::
234 AspIleLeuThrLeuLysThrAlaSerGluAsnSer... 246
    ::::::::::::::::::::
775 GACCCGCTACCCCTGCTGAAATGAGACGACGCGCCACCATGACGCC 824
    ::::::::::::::::::::
824 AasnLysGlyArgGluLeuThrPylLysGlyProSerLeuGlnArg 262
    ::::::::::::::::::::
825 CAAAGACGCTGAGGACCTGCTGACAGAAATTAACACCTTACATGCGGC 874
    ::::::::::::::::::::
875 AsnLeuSerGlnMetLeuAspIleValAsnThrLysSerAsn 276
    ::::::::::::::::::::
875 TGCAATGAACCTTAAAGCTGCGCAAGC...CTGAAGAACGAGAAAGC 921
    ::::::::::::::::::::
277 GlnArgGluHisGlnAsnIleGlnAspLeuGluIleGluAsnGlu 293
    ::::::::::::::::::::
922 CAGATGAT...CCCAAAATCAACACCTTGGAGAGGAATGAGACCT 968
    ::::::::::::::::::::
293 uLysGluAlaGluAsnArgLysIleGlnGlnGluAsnArgIleLeuLysAsp 310
    ::::::::::::::::::::
969 TTCTCTTAAGCTGCGGAGATTGCCAGATCACTGACGACGCTACAGATG 1018
    ::::::::::::::::::::
310 yValAsnGlyLeu... 314
    ::::::::::::::::::::
1019 CCTCATGATGCTGACGAGGACGACACGACGACCTGACGAGGCGCA 1068
    ::::::::::::::::::::
315 ...GlnLeuGlnLeuAsnGlnGluValMetValAlaAspPheLys 329
    ::::::::::::::::::::
1069 GAGAAAGACGCCGACGTGAAAGAGAGAGCTGACGACGCC...CTGCA 1112
    ::::::::::::::::::::
329 uSerGluLys...GluLysLeuLysSerLeuLeuAlaLysGlu 344
    ::::::::::::::::::::
1113 GAGCAAGAAATGCCCTTGAGAGAGAAACGAATCTTCAGGAAACTTT 1162
    ::::::::::::::::::::
344 yGlnIleGlnGluLysSerLeuAsnArgThrIleGlu... 354
    ::::::::::::::::::::
1163 CAGAGCTGAAAGAAACACTTCTCCACGCTCAGAGATACCCACCCGAG 1212
    ::::::::::::::::::::
354 ... 354
    ::::::::::::::::::::
1213 AAGGCGAGGTGCTGGGTGATGTCCTGACGTGAACCTTGAAGCA 1262
    ::::::::::::::::::::
355 ...AlaLeuLysSerArgPheLys 362
    ::::::::::::::::::::
1263 GCGAGCACTGCTGCTGCAACACACACAGCTCAGACGCGAGGAGAA 1312
    ::::::::::::::::::::
362 yPheGluSerAspHis...LeuLysSerGly 371
    ::::::::::::::::::::
1313 TGCTGAGACTGAGCGACGACGACGAGAACTGCTGCTGAGCGG 1362
    ::::::::::::::::::::
372 SerHisPheAsnGlyGluAspMetLeuLeuLysGln... 383
    ::::::::::::::::::::
1363 GCGCACTTGAAGAAAGAAAGACAGCTGCTGACGCTGATCACTAGCT 1412
    ::::::::::::::::::::
383 ... 383
    ::::::::::::::::::::
1413 GCGAGCTCATCTCAACCTCAGCAGGCGCAAGAGAGAGCTGACAG 1462
    ::::::::::::::::::::
384 ...GlyGlnMetTyrMetGlnAsp 390
    ::::::::::::::::::::
1463 CCTCCAGGCTCATAGGAGCCGCTTGAAGCTGACGAGTGGCTCTGAGC 1512
    ::::::::::::::::::::

```

```

391 SerGlnCysThrSerThrGlyMetProValHisMetGlnSerArgSer 407
    ::::::::::::::::::::
1513 TCTGAGCTACCAACCTCATAGTCAACATGACCATCAAGCAAGATCAAACT 1562
    ::::::::::::::::::::
407 LeuArgProLeuGlnIleLeuMetProValHisMetGlnSerArg 424
    ::::::::::::::::::::
1563 GCGTCCCTGAAGACAGCGCCAAAGGAAGCAAGCGCCACTGACGAA 1612
    ::::::::::::::::::::
424 sngIleIleLeuLysGlyGlnLeuGlnIleMetArgThrPheCysAspSer 440
    ::::::::::::::::::::
1613 CC...CTCCAAACGCAAGAACAGGCGCTCCAGGCGCTCGCCACG 1656
    ::::::::::::::::::::
441 AlaLysGlnAspArgLeuLysGlnIleAsnGlnIleValHisLysVal 457
    ::::::::::::::::::::
1657 GTGAGACAG...CTAAGCATGCTGCTGAAGCAAGAGCA 1694
    ::::::::::::::::::::
457 aGluCysLysAlaLeuAlaLeuGlnCysGlnArgValLysGluAspSer 474
    ::::::::::::::::::::
1695 GCACTGTAAGAGAGCTAGCGAGAGAGAGAGGACATGACGAGACCATG 1744
    ::::::::::::::::::::
474 sPcGln... 1744
    ::::::::::::::::::::
1745 CCGACACTGCGCACTGCTGCGAGAGAGACCAAGAGCGCTCTTAAGAG 1794
    ::::::::::::::::::::
480 LeuGluAspAlaLeuLysAspValGlnLysArgMetTyrGluSerGln 1956
    ::::::::::::::::::::
1795 CCGAGTACGCGCTCATACACAGCTGAGGAGCGCTGAGAAAGAGAGCTGC 1844
    ::::::::::::::::::::
496 yLysValLysGlnMetGlnThrHisPheLeuAlaLeuLysGlnIleLys 513
    ::::::::::::::::::::
1845 CAAGCTGAGATTCTGACGACCAACTGACAGTCTAAATGA...G 1888
    ::::::::::::::::::::
513 hSerAspAlaAlaThrGlyAsn... 520
    ::::::::::::::::::::
1889 CCGGACACAGTCCCAACACTTCAATGACACAGCGCCAGCGAGAGCA 1938
    ::::::::::::::::::::
521 ...HisArgLeuMetGlnGlnIleuLysAspGlnLeuLysAspMetLys 535
    ::::::::::::::::::::
1939 GAGCTGACCGGAAAGCTGAGGAATCTCAGGCTCTGTTGAAGCCG 1986
    ::::::::::::::::::::
535 sValLysTyrGluGlyAlaSerAlaGluValGlyLysLeuArgAsnGln 552
    ::::::::::::::::::::
1989 CCAAGAACAGATGAGCCCAAGCGCCAGCTGACACCTTGAAGCTCCAGC 2038
    ::::::::::::::::::::
552 LeuGlnAsnGlnMetLeuValGlnGluPheLysArg...Asp 565
    ::::::::::::::::::::
2039 TCCGCTGACGACCAAAAGCAACTGAAAGAAAGAGCGTGGCCAGAG 2088
    ::::::::::::::::::::
566 GlnGlyLysLeuMetGlnGlnLysLysArgGlnGlnGlnGlnLeuSer 582
    ::::::::::::::::::::
2089 AAGGACCAAGCTCCAGGAGACGCTCAGCGCCCTCAAAAGCTCTTGAAGCT 2138
    ::::::::::::::::::::
582 LcysGlu...LeuGluArgGluLysArgGluLysArgLysLeuThrGlu 597
    ::::::::::::::::::::
2139 CACCAAGGAGCGCTTGAAGAGAGAGAGC...AGCGCTCAGATGCC 2185
    ::::::::::::::::::::
597 eGlnGlnGlnLeuLysAspLeuSerAlaLysLeuAlaLeuSerIlePro 613
    ::::::::::::::::::::
2186 TGGAAGACGACGAGCGCTTATCTCT...GAGCTGAAG 2220
    ::::::::::::::::::::
2221 GCAGAG...ACCAAGACCTGAG...GAGCAAG 2252
    ::::::::::::::::::::
614 AlaGlnLysPheGluAsnMetLysSerLeuLeuSerAsnGlnLeuAsnGln 630
    ::::::::::::::::::::
630 uLysAlaLysLysLeuIleAspValGluArgGlnLysGlnLysSerLeu 647
    ::::::::::::::::::::
2253 GCGGAGAGGAAGAGCTGGAAGAGAGAGAGCGTGGCGCCAGAGCGCTG 2302
    ::::::::::::::::::::
647 sN...GluThrArgPro 651
    ::::::::::::::::::::
2303 AGGCTGATTAAGTGAAGCTTGGGAGAGCGCCATCAAGCTGACAGTCT 2352
    ::::::::::::::::::::

```

```

652 leuylarTgIuLeuGIuAnleuLySaLeIeuaLeuIaGlnIleVally 668
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2353 CTGGCGGGGGAGCTG.....CGAGGGGAGCTGGCTGGCGACACACAGC 2396
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
668 aProGIuIuIaGIuIeGIuIeLeuLySeTgIeU..... 679
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2397 TGAGAGTGAAGCTGGCTGCAAGAGAGAGTACCTGGCGGTGAGC 2446
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 .....GluGIuLySeTgIeUleuGIuLySaTgIeThr 691
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2447 GGTATGAGATGACGACGACAGAGAGCGACACACAGTGGCGATGGTACG 2496
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 GIuLeuThrSeTgIeUaNGInThrGIuIeGIuIeGIuIeGIuIeValy 708
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2497 GAA.....CACTGATGACTTTGACAGAGGAATGAGAGAGGCGC 2537
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 aLeuSaPaNulysleuThrGIuIeValSaMaNleuThrThrGIu 725
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2538 CAGAGAGCTGCGAGGCAAGAGAGAGTGGCGGCTGAGTACAGTCCACA 2587
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 eLySaMaNulProIuLySaISeTgIuIeGIuIeLySeThrSaP 741
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2588 GCGAG.....CTCCAGTACGCGCGGACGCGAGAA..... 2619
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 ValIleValSaPaNulSaNulysleuThrSaPValThrIleValy 758
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2620 .....CTAGCTGAGCTGACGACCTGGCGAGAGACGACGACGAGT 2663
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
758 ThrGIuIeLySaISeTgIuIeGIuIeGIuIeGIuIeGIuIeGIu 769
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2664 CCAAGACAGAGAGTCAAGAGGCGGCGGACCTTGCAGACCTCTCCATC 2713
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
769 eUmeGCIuAn..AlaSeTleuSeTlySaMaNulSeTATgIuGIuThr 784
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2714 TGCAGAAAGAGTGGCTGCACCCGACAAAGAGTGGCCCGCTTGAGAC 2763
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
785 ValPhe..... 786
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2764 TTGGTCCGACAGGAGTGAACGACGAAAGACCTCCCGGAGATTAAT 2813
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
787 .....IleProGIuIaThrGIuIeGIuIeGIuIeGIuIeGIu 795
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2814 CAAGAGCTCGGAGGAGAGACACACACACCCGAGTGGCTGGAAAGC 2863
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
795 Lu.....MeTAlaIleuLyS 800
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2864 AACAGGAGCGGAGCTTGCAGACACAGCGACGCGCTGACAGCTTGGAG 2913
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 SeTAlaIleThrGIuLeuLySaGIuIeSeTgIeU..... 813
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2914 CCGGAGGCGACAGATGGCCAGTGAAGTGGAGCGCTGGCGGCGCGCT 2953
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
814 .....AaLySaLySeGIuIeGIuIeGIuIeGIuIeGIuIeGIu 823
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2964 GATGAGAGCGAGGCGACGACGAGAGAGAGCGTGGACGAGAAAGG 3013
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
823 uLyIeThrSeTleuSeTlySaMaNul..... 832
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3014 AGCTGCGCGGCTGACCGAGAGAGCGGCGCTGGCCAGGCTGACCTGGC 3063
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
833 .....AaLySaLySeGIuIeGIuIeGIuIeGIuIeGIuIeGIu 838
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3064 CTGGAGAGCGGCGACAGACAGAGCTGATGAGCTGGCTGACGAAGCGGCT 3113
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
838 tSeThrIaGIuIeValIProValIySaThrIleGIuIeGIuIeGIuIeThr 855
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3114 CAGCGAGAGCGCTGGAGCTGCTACCGTCCAGAGAGAGCGAGCTGGCTATC 3163
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855 eUSeTSeTThrleuSaPulysThrSaMaNulGIuIeValSaPulys... 870
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3164 CCGTCAAGAGAAAGAGCAAGAGACAGAGATGGCAAGCTTGGCTGGCT 3213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
871 .....LySaLySeGIuIaPThrIleNGInGIuIePheVa 881
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

3214 CTGGAGAGCGGCGGACATGAAAGAGCTGAGAGGAATCTGCCAAGCGTGA 3263
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
881 llySaIleSaPaNulSaNulIleuIeLySaThrSaMaNulGIuAnThrG 898
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3264 GCAGCTGAAGGAA.....CACTGGCTAGAAAGAAAGAGAGAGCGGAT 3307
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
898 lSaNGIn..... 900
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3308 CTGGCTGAGAGCGCAATCTGAGAGCTGGCGAGAGACAGCAAGCGC 3357
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 .....VallySaIleGIuIeThrIleSeTleU..... 908
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3358 CCGCAAGCTGAGACAGCTGGCGGAGAGTGGAGAGCTGGACAGCATG 3407
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
909 ..ArgGIuIaGIuIeGIuIeSeTgIeUaIySaMaNulLySaP 925
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3408 CCGAGAGCGAGAGAGAGCTGACAGCGCTGGAAGAGAGCTGACAGCTG 3457
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
925 aGIuIaPaNulSaPulSaGIu.....IleGIuIa 934
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3458 AGCGGCGCTCGCGGCTGAGCGGAGAGCGAGCTGCTGAGAGCTCGAGG 3507
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
935 lYsTgIySaSeTgIuIeGIuIeValThrIleGIuIeGIuIeGIuIe 951
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3508 CAGTTAGAGAGAGAGCGGCGAGAGCTGAGAGAGAGAGAGCTTGGC 3557
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 aIaGIuIySaIySaIleuSaPThrIleGIuIeGIuIeGIuIeGIuIe 964
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3558 CTGGCGGCAAGGAGATTGGCTGCTTCCGACCAAGGTACAAAGACACA 3607
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
964 ..... 964
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3608 GCAGAGCTGAAGATGATGAGAGCGGAGCGGAGCTGGCGGCGGCGGCAAG 3657
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
965 ...LySaIleuLySaTgIaProIleIleSeTleuGIuIe..... 976
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3658 GCTGAGAGGAAAGATGAGCTTCACTGACGACTTGGAGAGAGAGCTGCTAT 3707
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
977 ..CySaIuATgIySaIleThrGIu.....LyGIuIeLyS 990
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3708 CCGAGATGCCAGAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3757
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 lUcIuIeSaNGInThrGIuIeLySaThrSaMaNulSeTgIuIeGIuIa 1006
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3758 GCGTGCATGCGCGAGTCAAGAAAGAGCGCAAGAGCTGGAGAGAGCTG 3807
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 lYsGIySaLySGInIuSaMaNulSaPulysleuLySaGIuIeIleThr 1023
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3808 GCGTGCATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3857
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 uGIuIe.....LySaPulysleu 1029
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3858 AGCTGCTGCTGGGAGAGAGCTGACAGAGCTGGCGGAGAGAGCTGATA 3907
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1029 sPulysaMaNulIleGIuIeSaNulSeTgIuThrGIuIaGIuIeSeT 1045
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3908 AACAGCGGAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3957
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1046 ArgIySaThrGIuIeSaMaNulGIuIeLySaPulysleuGIuIeGIuIe 1062
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3958 GAGCTGCGAGAGAGCTGGCCAGAGATGAGAGCGCTGGCGAGAGAGAT 4007
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1062 ThrGIuIaIySaGIuIeGIuIeGIuIeGIuIeGIuIeGIuIeGIuIe 1078
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4008 CTTCGAG 4057
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1078 yGIuIe.....ThrSeTgIeUaIleGIuIeValIleGIuIeGIuIeGIuIe 1091
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4058 CAGAGCGGCTGATGAGCTGCTGCGAGAGAGAGAGAGAGAGAGAGAG 4107
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1092 .....LyGIuIeValIProIeGIuIe 1099
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

4108 CTCAGAGCCGACAGAGCCGCTGCCGAGAAACCCAC.....CTGAGGA 4151
1099 nValGuserLeuLysSerLeuSerGlyThrIleGluThrLys 1116
4152 GCTGAGAGAGACAGAGAGCCGCTGGGGA.....CTGCGGG 4189
1116 lueGluLeuLysThrLysGlnArg..... 1123
4190 CAGAGCTGCTGGGCCCCAGGCGGAGCTGGGAGCTATTCCTGCGG 4239
1124 ..CysTyrGluGluGlnGlnThrValThrIleLeuArg..... 1136
4240 CAGAGGTGGCAGAGAGAGAGAGAGCTGACAGCTGGGCGAGAA 4289
1137 .....GlnMetLeuGluGlnGlnLysAsnSerValProL 1149
4290 GGCACGCTATGACAGCAGCTGAGCATCTCAAAAGGCGATGCGCTCG 4339
1149 euaLagLueLysLeu..... 1153
4340 TGCGCAGAGAGAACCGGGGCTGGGTGAGCGGCCACCTTGGCGGAG 4369
1154 .....GlnValLysGluAlaPheGluLysGluVal 1163
4390 TTCTGAGAGCTGAGGTGAGACAGAGCCGCAAAATATGTCAGAGCTT 4439
1163 lGylIleLysLysSer.....LeuArgLysGlnGluG 1176
4440 GGCAGCCGTGACCTGCTGATCTGAGACCGCTGCTGCTGAGTGCAGG 4489
1176 luserGlnAsnLysThrGluLysLysSerLysLeuGlnLysSerGlu 1192
4490 AGCAGAGAGCAGCTGGCGGAGCTGAGAGTGTGATGCTGCGAGATGAG 4539
1193 AsnThrLysGlnAlaLeuLysLysLeuGluThrArgGluValAlaPhe 1209
4540 GGTGGCAG.....GTCAAGCTCTGTGAGAGAGCGAG 4572
1209 userLysTyrLysAlaThrLysSerAspLeuGluThrGlnIleSerAspL 1226
4573 ....CGGTCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4618
1226 euaAsnGluLysLeuAlaAsnLeuAsnArgLysTyrGluGlnGluVal 1242
4619 TCGATGAGAAAGTGGCTGAC.....TCTGACCAAGCCAGC 4653
1243 GluValLeuAlaLysLysLysGluLysSerAla.....LysAs 1266
4654 AGGTGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4703
1266 pGluLysGluLeuLysPheSerIleGlu..... 1266
4704 CACAG 4753
1267 .....GlnGluLysAspArgIleGlu 1274
4754 AGTTGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4803
1275 ATGCGAspLysSerLeuThrIleThrGlnLeuGlnArgGluIle 1291
4804 AAGCCAAACATATTATGATCCCAAGAGAGAGAGAGAGAGAGAGAGAG 4853
1291 nGluSerAlaLysGlnIleGlu.....AlaLysAspLysIleThrG 1306
4854 GAGCAG 4897
1306 lueLeuAsnAspArgIleGluLysGlnAlaLeuAsnGluLysSer 1322
4898 AGCTGAG 4938
1323 GlnLeuThrTyrGlySerProSerLysArgGlnSerGlnLeuI 1339
4939 .....CAGCCTGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4976

```

```

seq_name: /gen2.6/ptodata/2/lna/5a.CONB.seq.us-08-467-781-3
seq_documentation_block:
  / Sequence 3, Application US/08467781
  / Patent No. 5780596
  / GENERAL INFORMATION:
  / APPLICANT: TONKATY, GARY
  / APPLICANT: LIDGARD, GRAHAM P
  / TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
  / NUMBER OF SEQUENCES: 6
  / CORRESPONDENCE ADDRESS:
  / ADDRESSER: TESTA, HUBERTZ & THIEDEULT
  / STREET: 125 HIGH STREET
  / CITY: BOSTON
  / STATE: MA
  / COUNTRY: USA
  / ZIP: 02110
  / COMPUTER READABLE FORM:
  / MEDIUM TYPE: PC compatible disk
  / OPERATING SYSTEM: PC-DOS/MS-DOS
  / SOFTWARE: Patentin Release #1.0, Version #1.25
  / CURRENT APPLICATION DATA:
  / APPLICATION NUMBER: US/08/467/781
  / FILING DATE: 06-JUN-1995
  / CLASSIFICATION: 424
  / ATTOBRY/ASB/PROTEIN INFORMATION:
  / NAME: PITCHER ESO, EDWARD R
  / REGISTRATION/DOCKET NUMBER: 27,829
  / TELECOMMUNICATION INFORMATION:
  / TELEPHONE: (617) 248-7100
  / TELEFAX: (617) 248-7100
  / INFORMATION FOR SEQ ID NO: 3:
  / SEQUENCE CHARACTERISTICS:
  / LENGTH: 4938 base pairs
  / TYPE: nucleic acid
  / STRANDEDNESS: single
  / TOPOLOGY: linear
  / MOLECULE TYPE: cDNA
  / FEATURE:
  / NAME/KEY: CDS
  / LOCATION: 1..6306
  / PUBLICATION INFORMATION:
  / JOURNAL: J. CELL BIOL.
  / AUTHORS: SZILAND, IVAN A
  / TITLES: PRIMARY STRUCTURE OF NUNA, AN INTRANSCULAR
  / TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
  / TITLE: SEGREGATION OF PROTEINS AT MITOSIS
  / JOURNAL: J. Cell Biol.
  / VOLUME: 116
  / PAGES: 1095-1408
  / DATES: 1992
  / US-08-467-781-3

```

```

Alignment_scores:
  Quality: 514.00      Length: 1541
  Ratio: 0.662        Gaps: 61
  Percent Similarity: 50.357      Percent Identity: 20.052

```

```

Alignment_block:

```

US-09-750-590-2 x US-08-467-781-3 ..  
 Align seg 1/1 to: US-08-467-781-3 from: 1 to: 6306  
 165 LysGlnAspArgThrAlaLeuMetLeuGlySerGlyTyrGlyValAspAsp 201  
 643 CAGATGAGCGGCTGAAAGAGACAGCTGCTGAGTACAGATAGAGTAAAGGGA 692  
 201 PALVALGLVAL... LeuLysAsnGlyValAspValThrLeuLeu 217  
 693 TGAAGTGAAGAGAGCTGACCTGACCAAGCCAGCTCTCCAGAGAGAG 742  
 217 SPALSLGGLYHisAspSerSerTyrTyrAlaHisGlyAspAspAsp 233  
 743 ATCCA..... CAGATGACCATGATGACAGAGCCGATTT 774  
 234 AspIleLeuThrLeuLeuLysThrAlaSerGluAsnSer..... 246  
 775 GACCCCTGACACCTGCTGATGATGAGAGAGCGGGCCAGCCACTGAGCC 824  
 247 AsnLysGlyArgGluLeuThrLysArgLysProSerLeuGlnArg 262  
 825 CAGAGAGCTTGAGAGAGCTGCTGACAGAGATAGAGAGCTTACATGAGGCG 874  
 263 ..... AsnLysSerGlnMetLeuAspLysThrLysSerAsp 276  
 875 TGCAATGAAGACCTGAGACAGCGCCAGAGC... CTGAGAGCAGAGAGAGC 921  
 277 GlnArgGlyHisGlnAsnIleGlnAspLeuGlyIleGlnAsnGluAspLeu 293  
 922 CAGATGAT... CGCAATTCAGCAGCTTCCGAGAGAGATGACAGAGCT 968  
 293 uLysGlnArgLeuArgLysIleGlnGlnGluIleLeuLeuAspL 310  
 969 TTCCTTAACTTCGGAGAGTTGCCAGTATCTCCAGAGAGCTACAGAGAT 1018  
 310 YValAlaMetGlyLeu..... 314  
 1019 CCTCATGAGCTGAGCGAGAGAGACAGCAAGCCACTGACAGTGGCTA 1068  
 315 ..... GlnLeuGlnLeuAsnGlnGluValAlaAspAspLeuG 329  
 1069 GAGAGACAGCCAGCAGCTGAGAGAGAGAGCTACAGCAGCC..... CTGCA 1112  
 329 uSerGluLys... GlnLysLeuLysSerLeuLeuAlaIleGluL 344  
 1113 GCAAGAGAAATGCTGTGAGAGAGAGAGAAATCTTCAGAGAAATCTTT 1162  
 344 YGlnIleGlnGluLysLeuArgThrIleGln..... 354  
 1163 CAGAGCTGAGAGAGACCTGCTCCAGCTGAGAGATAGCAAGCCAGAGAG 1212  
 354 ..... 354  
 1213 AAGGGAGAGCTGCTGGTATGATCTTGACAGCTGAGAACTTGAAAGAGA 1262  
 355 ..... AlaLeuLysSerArgPheLeuArg 362  
 1263 GGCAGGCACCTTCTGCTCAACACAGACAGCTCCAGCCAGAGGTGAGA 1312  
 362 YPheGluSerAspHis..... LeuGlySerGly 371  
 1313 TGTGTGAGACTGTAGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1362  
 372 SerHisPheArgLysGluAspPheLeuLeuLysGln..... 383  
 1363 GGGCACTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1412  
 383 ..... 383  
 1413 GCAAGAGCTCAATCTCAACCTCAAGCAGCCAGAGAGAGAGAGAGAGAG 1462  
 384 ..... GlyIleMetLysThrIleAsp 390

1463 CTTCCAGAGCTCAATGAGGCGCGCTTGGCTGAGAGAGAGAGAGAGAG 1512  
 391 SerGlnCysThrSerThrGlyMetProValHisMetGlnSerArgSerLeu 407  
 1513 TGTGAGCTCAACACCTCAATGACCCAGCTCAGCAAGAGATCAAGCAACT 1562  
 407 LeuArgProLeuGlnLeuAlaLeuProAsnGlnAlaSerTyrSerGln 424  
 1563 GGTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1612  
 424 AsnGluIleLeuLysArgLeuGlnAlaLeuLysArgPheLysAspSer 440  
 1613 CC..... CTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656  
 441 AlaLysGlnAspArgLeuLysLeuGlnAsnGlnLeuAlaHisLysVal 457  
 1657 GTGGAGAGC..... CTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694  
 457 AsnLysGlyAlaLeuAlaLeuGlnLysGlnArgValLysGlnAspSer 474  
 1695 GCGATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1744  
 474 SPGLuGln..... IleLysGln 479  
 1745 CCGAGACAGCTGGCCAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1794  
 480 LeuGlnAspAlaLeuLysAspValGlnLysArgMetTyrGlnSerGlu 496  
 1795 CGGATCGGCTCTCAAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1844  
 496 YLysValLysGlnMetGlnThrHisPheLeuAlaLeuLysGlnIleLeu 513  
 1845 CAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888  
 513 HisSerAspAlaIleThrGlyAsn..... 520  
 1889 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938  
 521 ..... HisArgLeuMetGlnGlnLeuLysAspIleLeuLysAspMet 535  
 1939 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1988  
 535 sValLysArgLysGlnAlaSerLeuGlnValGlyLysLeuArgAsnIle 552  
 1989 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038  
 552 LeuLysGlnAsnGlnLeuValGlnIleLysArg..... Asp 565  
 2039 TGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088  
 566 GluGlyLysLeuMetGlnGlnLysLeuArgLeuIleLysGlnLysLeu 582  
 2089 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2138  
 582 LysGln..... LeuGlnArgLysArgGlyArgLysLeuThrGln 597  
 2139 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2185  
 597 AsnLysGlnLysLeuLysAspLeuAlaLysLeuAlaLeuSerIlePro 613  
 2186 TCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2230  
 614 AlaGlnArgPheGlnAsnMetLysSerLeuLysSerAsnGlnLeuAsnG 630  
 2221 GCAGAG..... ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 647  
 630 uLysAlaLysLysLeuIleAspValGlnLysGlnLysLysSerLeu 647  
 2253 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2302  
 647 sn..... GluThrArgPro 651

2303 AGGCTGATTAATGACCTTGGGAGGCGCCATGACAGCTGACATCTGAC 2352  
652 LeuLysAlaArgIleLeuGluLysAlaLysLeuAlaGlnIleValTyr 668  
2353 CTGGGGCGGGAGAGCTG ..... GCGAGGCCATGAGCTGCCCGACACACAC 2396  
668 sProGluGlnHisGluIleLeuLysSerArgLeu ..... 679  
2397 TGAGAGTATGATGTGACAGCTCTCCAGAGAGAGAGAGAGAGAGAGAGAG 2446  
680 ..... GluGlnLysSerGluLysLeuGluLysGluLysLeuHis 691  
2447 GGTATGAGATAGCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496  
692 GluLeuHisSerLysAsnGlnIleLeuGluLysGluLysGluLysValTyr 708  
2497 GAA ..... CAGCTGATCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2537  
708 sLeuAspAsnLysLeuLeuThrGlnGlnValAlaAsnAlaLeuThrGln 725  
2538 CCGAGAGCTCCAGAGGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2587  
725 cLysAsnValIleProLeuLysValSerGluIleLysLysSerHisAsp 741  
2588 GCGAG ..... CTCCAGATAGCGCGGAGCGACACCAACAA ..... 2619  
742 ValIleValAspAsnLysLysLysSerAspValThrIleLysTyr 758  
2620 ..... CTAGCTGAGCTCCATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2663  
758 ThrGlnLysLysLeuGluIleMetLysLeu ..... L 769  
2664 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2713  
769 eLysGluAsn ..... AlaSerLysSerLysAsnValSerThrGluGlnThr 784  
2714 TCGAGAGAAAGATGGCTCGCCACCGCAAAAGAGAGAGAGAGAGAGAGAG 2763  
785 ValPhe ..... 786  
2764 TTGGTCCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2813  
787 ..... IleProGluLysArgHisGluLysG 795  
2814 CAGAGAGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2863  
795 Lu ..... MetMetAlaLeuLys 800  
2864 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2913  
801 SerAsnIleThrIleLysLysGlnLeuSerGluLeu ..... 813  
2914 CGGGAGCCAGAGAGATGGCAATAGCTGAGAGAGAGAGAGAGAGAGAGAG 2963  
814 ..... AsnLysLysCysGlyGluAspGlnIle 823  
2964 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3013  
823 ValLeuTyrSerLeuMetSerGluAsnAsn ..... 832  
3014 AGGTGGCGCGGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3063  
833 ..... AspLeuLysLysPhe 838  
3064 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3113  
838 sThrIleGluTyrValProValLysThrHisGluGluIleLysThrAla 855  
3114 CACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3163  
855 eLysSerThrLeuAspLysThrAsnArgLysGluValValAspValLys ... 870  
3164 CCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3213

---

871 ..... LysLysCysGluAspIleAsnGlnIlePheAs 881  
3214 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3263  
881 LysIleLysAspIleAsnGlnIleLysLysArgLysGlnIleAsnHis 898  
3264 GCAAGTAAAGGAA ..... CAGCTGGTAAAGAGAGAGAGAGAGAGAGAG 3307  
898 LysAsnGln ..... 900  
3308 CTGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3357  
901 ..... ValValAlaGluTyrIleSerLeu ..... 908  
3358 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3407  
909 ArgGlnHisGlnIleLysMetSerGluLeuArgLysSerMetLysLysVal 925  
3408 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3457  
925 AlaGlnAspSerAlaGlu ..... IleLeuAla 934  
3458 AGCGAGCTCCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3507  
935 LysTyrLysSerGlnGlnIleValThrLysGlnGlnIleAla 951  
3508 CAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3557  
951 AlaGlnIleValArgIleLysAspThrIleGlnLysTyr ..... 964  
3558 CTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3607  
964 ..... 964  
3608 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3657  
965 ... LysLeuLysTyrAlaProGluIleSerLeuGlnGlu ..... 976  
3658 CCGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3707  
977 CysGluAspLysPheLysAlaThrGlu ..... LysGlnLysLysG 990  
3708 CCGTAATCCAGAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3757  
990 LysIleLysSerGlnIleThrGlnLysTyrAsnThrSerGlnGlnAla 1006  
1007 LysLysCysArgLysGlnLysAspLysLysLysLysLysLysLysLeu 1023  
3808 GCGTGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3857  
1023 uGln ..... LysAspLeuLysA 1029  
3858 AGCTGTCTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3907  
1029 sPheLysAsnValHisGluLysAsnSerTyrGlnHisGlnArgLysAsp 1045  
3908 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3957  
1046 ArgLysThrGlnLysLeuAsnArgLysLysLysLysLysLysLysLys 1062  
3958 GAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4007  
1062 ThrGlnAlaLysLysGlnLysLysLysLysLysLysLysLysLysLys 1078  
4008 CTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4057  
1078 uArgLys ..... ThrSerGluIleLeuAlaAlaGlnIleLeuLysGln ... 1091  
4058 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4107

```

4939 .....CAGCGTCGGCGCGAAGACAGAGAGCGTCAAGCGCTG 4976
1339 eapserfenglnclnglnvlnvlnvserfenglnclnglnleahlaapa 1356
4977 CGCCACCTACTCTCCCAAGGCGCGCGACGCTCGAGCGACGAGCGTTCGATG 5026
1356 laapafgcnlnhlsnglnclnglnvln 1363
|||||.....
5027 CAGACCGACGACGCTTCGAGCGCTG 5049
seq_name: /cgn2.6/prodata/2/lna/5b_comp.seq:US-08-483-924-3

seq:documentation_block:
1 seq:us08483924 location US/08483924
2 Patient No 5882876
3
4 GENERAL INFORMATION:
5 APPLICANT: TOKUAKI, GARY
6 APPLICANT: LINDARD, GRAHAM P
7 TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF T
8 TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
9 NUMBER OF SEQUENCES: 6
10 CORRESPONDENCE ADDRESSES:
11 STREET: 125 HIGH STREET
12 CITY: BOSTON
13 STATE: MA
14 COUNTRY: USA
15 ZIP: 02110
16
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMMENTS: IBM PC compatible
20 SOFTWARE: Patent in release #1.0, Version #1.25
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/483,924
24 FILING DATE: 07-JUN-1995
25 CLASSIFICATION: 435
26
27 ATTORNEY/AGENT INFORMATION:
28 NAME: PITCHER SSO, EDMUND R
29 REGISTRATION NUMBER: 27,829-P-013
30 REFERENCE/DOCID INFORMATION:
31 TRADEMARK INFORMATION:
32 TELEPHONE: (617) 248-7000
33 TELEFAX: (617) 248-7100
34 INFORMATION FOR SEQ ID NO: 3:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 6306 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 FEATURES:
41 FEATURE TYPE: CDS
42 NAME/KEY: CDS
43 LOCATION: 1..6306
44 PUBLICATION INFORMATION:
45 AUTHORS: COMPTON, DUANE A
46 AUTHORS: SZILAK, ILVA
47 AUTHORS: CLEVELAND, DON W
48 TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
49 TITLE: PROTEIN THAT DEFINES A NOVEL SUBSTRATE FOR
50 TITLE: SEQUESTRATION OF PROTEINS AT MITOSIS
51
52 VOLUME: 116
53 PAGES: 1395-1408
54
55 DATE: 1995
56
57 US-08-483-924-3
58
59 Alignment_scores:
60 Quality: 514.00 Length: 1541
61 Ratio: 0.652 Gaps: 671
62 Percent Similarity: 50.357 Percent Identity: 20.052

```



```

alignment_block:
US-09-750-590-2 x US-08-483-924-3 ..
Align seg 1/1 to: US-08-483-924-3 from: 1 to: 6306

185 LysGIlnAspArthrAlaLeuLeuLeuGlyCysGlyTrpGlyCysLysAs 201
      ::::::::::::::::::::
643 CAGATGAGAGCGCTGAGAGACAGACGCTTGCTGATGAGAAAGATGATGAG 692
      ::::::::::::::::::::
201 PALAVALGVAL...LeuIleLysAsnGlyAlaAspAlaThrLeuLeuA 217
      ::::::::::::::::::::
693 TGAGCTGAGCTGAGACCTGAGAAACCGACGACCTTACCGAGAG 742
      ::::::::::::::::::::
217 SPALaLeuGlyHisAspSerSetrTyTrpAlaArGlyLysAspLeu 233
      ::::::::::::::::::::
743 ATCCA.....CAATAGCCATGATGCTCAAGACGCGAT 774
      ::::::::::::::::::::
234 AspIleLeuThrIleLeuLysThrAlaSerGIlnAspSer..... 246
      ::::::::::::::::::::
775 GACCGCTAGCCCTGCAATATGACAGACGAGCGGCGACCCGACGAGCC 824
      ::::::::::::::::::::
247 AsnLysGlyArGlyLeuLeuTrpLysGlySyrProSerGlyGlnIleArG 262
      ::::::::::::::::::::
825 CAAGAGACCTGAGAGAGCGCTGACAGAAAGAACGCTTACCGATGGCC 874
      ::::::::::::::::::::
263 .....AsnLeuSerGIlnMetLeuAspGlyAlaSerTrpLysSerAsn 276
      ::::::::::::::::::::
875 TGCAAGAAACCCCTGAGACGACTCCAGAGC...CTGAAAGACAGAGAGC 921
      ::::::::::::::::::::
277 GlnArGlyGlnHisGlnAsnIleGlnAspLeuGlyIleGlnAspGlnAsp 293
      ::::::::::::::::::::
922 CAAGATGAGAT...CCGAAATATCAACCAAGCTTTCGAGAGAAATGAGACCT 968
      ::::::::::::::::::::
293 LysGlyLysArGlyLeuArGlyTrpIleGlnGlnGlnLysArGlyIleLeu 310
      ::::::::::::::::::::
969 TTCCTTTAAGCTCCGGAGATTGCCAGTCATCTGACGACAGCTAACAGATG 1018
      ::::::::::::::::::::
310 LysValAspGlyLeu..... 314
      ::::::::::::::::::::
1019 CCTCATATGACCTGACAGAGACAGACAGACGACCTGAGAGTGCCTA 1068
      ::::::::::::::::::::
315 .....GlnLeuGlnLeuAsnGlnGlyValLysLeuAlaAspAspLeuGI 129
      ::::::::::::::::::::
1069 GAGAGACAGACCCGACCTGAGAGAGAGACCTCAAGCCAGCC...CTCA 1112
      ::::::::::::::::::::
329 userGIlys.....GlyLysLeuLysSerLeuLeuAlaLeuLysGlnL 344
      ::::::::::::::::::::
1113 GGCAGAAAGAAATGCTTTCAGAGAAAGAACCAAAATCCTTTCAGGGAAA 1162
      ::::::::::::::::::::
344 LysGlnHisGlnGlnLysSerLeuArGlyThrIleGln..... 354
      ::::::::::::::::::::
1163 CACAGCTGAAAGAACACTTGTCCACGCTGACAGATTAACCAACCCAGAG 1212
      ::::::::::::::::::::
354 ..... 354
      ::::::::::::::::::::
1213 AAGGAGAGGTGCTGGGTGATGTCTTGACAGCTGAAACCTTGAAGCAGA 1262
      ::::::::::::::::::::
355 .....AlaLeuLysSerArGlyPheLysTrp 362
      ::::::::::::::::::::
1263 GCGAGCGACCTTGCTGCTCAAAACAAACACAGAGCTCAAGAGAGCTGAGA 1312
      ::::::::::::::::::::
362 YrPheGlyUserAspHis.....LeuGlySerGly 371
      ::::::::::::::::::::
1313 TGTGAGACACTGAGCGAGCCACAGACAGAACCAAGCTGTGTGTCGACGG 1362
      ::::::::::::::::::::
372 SerHisLysPheArGlyGlyLysPheLeuLeuLysGln..... 383
      ::::::::::::::::::::
1363 GGCACCTTCGAGAAAGAAAGACGACGAGCTGTTCAGCCGATCAACAGAC 1412
      ::::::::::::::::::::
383 ..... 383
      ::::::::::::::::::::
1413 GCAGAGCTCCATCTCAACCTGACGACGAGCGAAAGAGAGCTGAGACAG 1462
      ::::::::::::::::::::
      ::::::::::::::::::::
384 .....GlyIleMethTrpMetHisAsp 390
      ::::::::::::::::::::
1463 CCTCCAGGCTCATGGGGCCGGTGAATGACTGACAGAGAGCTGTCTAGAC 1512
      ::::::::::::::::::::
391 SerGIlnCysThrSerTrpGlyLysProValHisMetIleSerAspMet 407
      ::::::::::::::::::::
1513 TGTAGCTCAACAGACATGATCCACATCCACAGACAGATCAAGACT 1562
      ::::::::::::::::::::
407 LLeuArGTrpLeuGlyLeuAlaLeuPheAsnGlnAlaSerTrpSerGlnA 424
      ::::::::::::::::::::
1563 GAGCTGGCTGAGACAGAGCGCCAAAGAGAGAGAGCCCAAGCTACAGCA 1612
      ::::::::::::::::::::
424 snGlnLLeuLysGlyGlnLeuGlnLysLeuAlaThrTrpPheCysAspSer 440
      ::::::::::::::::::::
1613 CC.....CTCCAAACAGCAAGAACAGGCTCCAGAGGCTCCGACAGAG 1656
      ::::::::::::::::::::
441 AlAlaLysGlnPArGlyLeuLysLysGlnAspGlyLysLeuValHisTrp 457
      ::::::::::::::::::::
1657 GTrGAGCCAG.....CTAAAGCTAGTCCGAGACCCAGAAAGAGACCA 1694
      ::::::::::::::::::::
457 nGlyCysLysAlaLeuAlaLeuGlyCysGlnArGlyLysGlnAspSerA 1744
      ::::::::::::::::::::
1695 CGAGTTGAAAGAGAGTACCGAGAAAGAGAGAGCAATCAGAGAGACATG 1744
      ::::::::::::::::::::
474 spGIlnGln.....IleLysGln 479
      ::::::::::::::::::::
1745 CCCAGCAACTGAGCCAGCTGCGAGAGAGAGAGAGAGGCTCTTAAAGGAG 1794
      ::::::::::::::::::::
480 LeuGlnAspAlaLeuLysAspValGlnLysArGlyMetTrpGlySerGln 196
      ::::::::::::::::::::
1795 CGGATGCGGCTCTCAAGCACTGAGAGACCTGAGAGAGAGAGAGAGAGCTG 1844
      ::::::::::::::::::::
496 LysValLysGlnMetGlnTrpHisPheLeuAlaLeuLysGlnHisLeuTr 513
      ::::::::::::::::::::
1845 CAAGCTGAGATTCGACAGACAGCAACTTCAAGTGTGATTAAGA.....G 1888
      ::::::::::::::::::::
513 HisSerAspAlaAlaThrGlyAsn..... 520
      ::::::::::::::::::::
1889 CCCGGAGACTGCCAGACCTCAATGACATCAAGACAGCCCAAGCGAGAGACA 1938
      ::::::::::::::::::::
521 .....HisArGLeuMetGlnGlyLeuLysAspGlnLysLysPheLys 535
      ::::::::::::::::::::
1939 GAGCTGAGCCGAGAGTGTGAGAGAGACTCAAGCGCTGTGTGAGACGCCG 1988
      ::::::::::::::::::::
535 SPALysTrpGlyGlnAlaSerAlaGlyValGlyLysLeuArGlnGlnI 552
      ::::::::::::::::::::
1989 CAGAGAACATGAGCCAGCCAGCCAGAGCTTCCAGAGAGACAGCTTGCGA 2038
      ::::::::::::::::::::
552 LysGlnAsnGlnLysLeuValGlnGlyPheLysArG.....Asp 565
      ::::::::::::::::::::
2039 TGCAGCTGACAGCAAAAGAAACAGCACTGAGAAAGAAAGAGGTGCCAGAG 2088
      ::::::::::::::::::::
566 GlnGlyLysLeuMetGlnGlnAsnLysArGlnGlnLysGlnLysSerMet 582
      ::::::::::::::::::::
2089 AAGGACAGAGCTCAAGAGAGAGCTCCAGAGCCCTCAAGAGAGCTTCAAGAT 2138
      ::::::::::::::::::::
582 LysGln.....LeuGlnArGTrpLysArGlyArGlyLysLeuThrGly 597
      ::::::::::::::::::::
2139 CACCAAGCGACCTTGAGAGAGAGAGAGAGCG...AGCGCTGCGATGCC 2185
      ::::::::::::::::::::
597 eGlnGlyGlnLeuLysAspLeuSerAlaLysLeuAlaLeuSerIlePro 613
      ::::::::::::::::::::
2186 TGAAGAGACAGACAGCTGTGTCTCT.....GAGCTAAG 2220
      ::::::::::::::::::::
614 AlAGlnLysPheGlnLysMetLysSerLeuLeuSerArGlnLysLeuAsn 630
      ::::::::::::::::::::
2221 GCAGAG.....ACCGAGAGCTGTGG.....GAGACACATTA 2252
      ::::::::::::::::::::
630 LysValLysLysLeuIleAspValGlnArGTrpGlyLysGlnLysSerLeuA 647
      ::::::::::::::::::::
2253 GCGGAGACAGAGAGAGACTGAGAGAGAGAGAGAGGCTGGCGCAAGAGGCTGG 2302
      ::::::::::::::::::::

```

```

647 80.....GluThrArgPro 651
2303 AGCTCGAATTATCCGCACTTGGGAGAGCCCTCAGAGCTCGAGACGCTGAGACGCTG 2352
652 LeuLysArgGluLeuGluLeuAlaLeuLysAlaLysLeuAlaGlnGlnVal 668
2353 CTGGGCGGGGAGCTG.....CCAGAGCGCTGCGTGGCCAGCAGCAGAC 2396
668 AProGluGlnHisGluGlnLeuLysSerArgLeu..... 679
2397 TGAAGAGTGGTGGACAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2446
2447 GGTATGAGATATGCGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496
692 GluLeuThrSerLysAsnGlnThrLeuGlnLysGluLeuVal 708
2497 GAA.....AGCTGATGACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2537
708 AlaAspArgLysLeuLeuThrGlnGlnLysAlaAsnAlaLeuThrThrLys 725
2538 CCAGAGCTCCGACAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2587
725 GelyAsnAlaProLeuLysValSerGluGluLeuLysLysSerHisAsp 741
2588 GCGAG.....CTCCAGATTAAGCCGCGCAGCAGAGAGAA..... 2619
742 ValIleValAspArgLeuLeuLysLeuSerArgLysThrHisLysArg 758
2620 .....CTAGCTGACCTTCATCCGACAGCTGCGCAGAGACCTCCAGCAGCT 2663
758 rThiLysLysLysLeuGluLeuLeuLysLeu.....L 769
2664 CCAG 2713
769 GluMetCysLysN.....AlaSerLeuSerLysAsnLysArgLysLeuGlnLysThr 784
2714 TCGAGGAAAGATGGCTGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2763
785 ValPhe..... 786
2764 TTGTCCGCAAGCCAGTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2813
787 .....IleProTrpGluArgHisGlyLysG 795
2814 CAGAGAGCTGGAG 2863
795 Lu.....MetMetAlaLeuLys 800
2864 AACAG 2913
801 SerAsnIleThrGluLeuLysGlnLeuSerGluLeu..... 813
2914 CCGAG 2963
814 .....AsnLysLysGlyLysAlaArgGlnGly 823
2964 GATGAG 3013
823 yAlIleTyrSerIleMetSerGluAsnAsn..... 832
3014 AGGTGGCTGGCTGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3063
833 .....AspLysLysLysThrPhe 838
3064 CTGGAG 3113
838 tSerHisArgLysValProValLysThrHisGluGlnLysLeuThrHisAla 855
3114 CAG 3163
855 eluSerSerThrIleuAspLysThrAsnArgGluLeuValAspValLys..... 870
3164 CCGTACAGCGAAAG 3213
871 .....LysLysGlyGluAspLysArgLysGlnGlnPheVal 881
3214 CTGGAG 3263
881 LlysLysLysArgLysAsnGlnLysLysLysLysLysLysLysLysLysLysLys 898
3264 GCACTGAGAGAA.....CACTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3307
898 LAsnGln..... 900
3308 CTGGCTCAGAGAGCCCATCTGAGAGCTGTGGCAGAGAGAGAGAGAGAGAGAGAG 3357
901 .....ValLysAlaGluLysLysLysLysLysLysLysLysLysLysLysLys 908
3358 CCGAGCTGGAG 3407
909 ArgGlnHisGluLysLysSerGluLysAlaGlyLysSerMetLysLysVal 925
3408 CCAG 3457
925 aLginAspAsnSerAlaGlu.....IleAlaAla 934
3458 AGCGAGCTCCCGGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3507
935 LysArgLysLysSerGlnGlnGlnLysLysLysLysLysLysLysLysLysLys 951
3508 CAGTTAG 3557
951 aAlaGlnLysArgGluLeuAspThrIleGlnGlyLys..... 964
3558 CTGGAGCCAG 3607
964 ..... 964
3608 GCAAGCTGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3657
965 .....LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 976
3658 GCTGAG 3707
977 CysGluArgLysArgPheLysAlaThrGlu.....LysGluLysLysG 990
3708 CCTGATCCCGCAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3757
990 LysGlnLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1006
3758 GCGTGTGTGATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3807
1007 LysLysCysLysGlnGlnLysAsnArgLysLysLysLysLysLysLysLysLys 1023
3808 GCGTGTGTGAG 3857
1023 uGln.....LysAspLysLysAla 1029
3858 AGCTGTCTCTGCGAG 3907
1029 sPyLysAsnValHisIleGluAsnSerTyrGluThrGluArgLysLeuSer 1045
3908 AACAG 3957
1046 ArgLysThrGluGlnLysLeuAsnArgLysLysLysLysLysLysLysLysLys 1062
3958 GAGCTGTGTGAG 4007
1062 rThiAlaLysLysGlnLysGluLysLeu.....ValGluAsnAlaL 1078
4008 CTTCGAG 4057
1078 yGln.....ThrSerGluIleLeuAlaGlnThrLeuGlnIn... 1091

```



```

alignment_scores:
  Quality: 512.00      Length: 1541
  Ratio: 0.660        Gaps: 61
  Percent Similarity: 50.357   Percent Identity: 20.052

alignment_block:
  US-09-750-590-2 x US-08-195-487-3 ..

Align seq 1/1 to: US-08-195-487-3 from: 1 to: 6306

185  LysGlnAmaAqPthAlaLeuMeuLgUyGyLurYqYyLys 201
    :::::::::::::::::::: ::::::::::::::::::::
643  CAGATGACGGCTGAGAACAGCAGCTTGCTGAGAAAGATAGGGA 692
    :::::::::::::::::::: ::::::::::::::::::::
201  pAlaValGluVal...LeuLleLysAnG1ValAspValTrhLeuLeu 217
    :::::::::::::::::::: ::::::::::::::::::::
693  TGAGCTGGAGCTGAGACAGTGGAAAGCCGAGCTCTCCAGCAAGG 742
    :::::::::::::::::::: ::::::::::::::::::::
217  gPAlaLleuG1LysAspSerSerTyTTrYAlaAlr1LeG1AspValLeu 233
    :::::::::::::::::::: ::::::::::::::::::::
743  ATGCA.....CAGATGACCATGATGACAGACGCGATT 774
    :::::::::::::::::::: ::::::::::::::::::::
234  AsP1LeuTrhLeuLleuLysTrhAlaSerG1LysAspSer 246
    :::::::::::::::::::: ::::::::::::::::::::
775  GAGCGCCCTGAGCTGCTGATAGAACAGCGCGCCAGCCAGCTGAGCC 824
    :::::::::::::::::::: ::::::::::::::::::::
247  AAnLysG1LysTrhLeuTrhPylYsLysLysProSerLeuG1NhrY 262
    :::::::::::::::::::: ::::::::::::::::::::
825  CAGAGAGCTTGAGAGCTGCTGACAGAGATGAGACCTTACCTAGCCG 874
    :::::::::::::::::::: ::::::::::::::::::::
263  .....AAnLysSerG1MeLLeuAspG1LysAlaTrhLysSerAsn 276
    :::::::::::::::::::: ::::::::::::::::::::
875  TGCGATGAACCCCTGAAGCATGCGAGAC...CTGAAGACAGAGAACG 921
    :::::::::::::::::::: ::::::::::::::::::::
277  G1AATG1LysL1SG1AAnG1LeG1AspValLeuG1LeG1LysAspLe 293
    :::::::::::::::::::: ::::::::::::::::::::
922  CAGATGAT...CCCAATCTAACAGCTTCTGGAGAGAAATGAGACGT 968
    :::::::::::::::::::: ::::::::::::::::::::
293  uLysG1LysG1LysL1LeG1NhrG1LysG1LysG1LysL1LeuLys 310
    :::::::::::::::::::: ::::::::::::::::::::
969  TTCTTTAAGCTGGCGGAGTTTCCAGCTCATCTGACAGACGTAAGCATG 1018
    :::::::::::::::::::: ::::::::::::::::::::
310  yAlaValAspG1Lys..... 314
    :::::::::::::::::::: ::::::::::::::::::::
1019  CCGCTAATGAGCTGAGAGAGAGACAGACAGACGCTACAGAGGCTA 1068
    :::::::::::::::::::: ::::::::::::::::::::
315  .....G1NhrG1LysAsnG1LysL1LysL1MeLValAlaLysAspLeuG1 329
    :::::::::::::::::::: ::::::::::::::::::::
1069  GAGAACGAGCGCCGAGCTGGAGAGAGAGAGCTACAGCGCC.....CTGCA 1112
    :::::::::::::::::::: ::::::::::::::::::::
329  uSerG1Lys.....G1LysL1LysL1LysL1LeuLysL1LysL1Lys 344
    :::::::::::::::::::: ::::::::::::::::::::
1113  GAGCAAGAAATGCTCTGAGAGAGAAAGCAAAATCTTCCAGGAAACTT 1162
    :::::::::::::::::::: ::::::::::::::::::::
344  yAG1NhrLysG1LysL1LysL1LysL1LysL1LysL1LysL1Lys 354
    :::::::::::::::::::: ::::::::::::::::::::
1163  CACAGCTGGAGAACACTTGTCCAGCTGCGAGATACCCAGCCAGAG 1212
    :::::::::::::::::::: ::::::::::::::::::::
354  ..... 354
    :::::::::::::::::::: ::::::::::::::::::::
1213  AAGCGGAGAGCTGCTGGCTGATGCTTCCAGCTGAGAACCTTGAAGACA 1262
    :::::::::::::::::::: ::::::::::::::::::::
355  .....AlaLysLysSerTrhPheLysTr 362
    :::::::::::::::::::: ::::::::::::::::::::
1263  GCGAGCCAGCTCTGCTGCAACACAGACGCTCAAGAGAGCTTGAAGA 1312
    :::::::::::::::::::: ::::::::::::::::::::
362  yTrhG1LysTrhAspLys.....LeuG1LysTrhY 371
    :::::::::::::::::::: ::::::::::::::::::::
1313  TGCTGGAGAGCTGAGAGAGAGCCAGAGAACAGACGCTGCTGCTAGCGG 1362
    :::::::::::::::::::: ::::::::::::::::::::
372  SerTrhLysPheTrhLysG1LysAspTrhLeuLysGln..... 383
    :::::::::::::::::::: ::::::::::::::::::::

```

```

1363  GCGACCTTGGAGAAAGAAAGACAGACAGCTGTACCTGATGACGCT 1412
    :::::::::::::::::::: ::::::::::::::::::::
383  ..... 383
    :::::::::::::::::::: ::::::::::::::::::::
1413  GAGAGAGCTCATGCTCCAGCTTCCAGCCAGCCAGAGAGAGCTGAGACG 1462
    :::::::::::::::::::: ::::::::::::::::::::
384  .....G1LysL1LysTrhTrhAsp 390
    :::::::::::::::::::: ::::::::::::::::::::
1463  CCGTCCAGGCTCATGAGGCGCGCTTACCTGCGAGAGCGCTCTGAGAC 1512
    :::::::::::::::::::: ::::::::::::::::::::
391  SerG1LysTrhSerTrhG1LysPheProValLysMeLTrhLysAspSer 407
    :::::::::::::::::::: ::::::::::::::::::::
1513  TCTGAGCTTCCAGCACTCAATCCAGCTACAGCAAGACAGCAAGACCT 1562
    :::::::::::::::::::: ::::::::::::::::::::
407  LLeuAspTrhLeuLysL1LysL1LysL1LysL1LysL1LysL1Lys 424
    :::::::::::::::::::: ::::::::::::::::::::
1563  GCGTCCGCTGAGAGAGAGAGCAAGAAAGAGAGAGAGAGAGAGAGAG 1612
    :::::::::::::::::::: ::::::::::::::::::::
424  snG1LleLysLysG1LysL1LysL1LysL1LysL1LysL1LysL1Lys 440
    :::::::::::::::::::: ::::::::::::::::::::
1613  CC.....CTCCAGACAGAAAGAGAGAGCTTCCAGAGAGCTTCCGAGCA 1656
    :::::::::::::::::::: ::::::::::::::::::::
441  AlAlysG1LysAspTrhLysL1LysL1LysL1LysL1LysL1LysL1 457
    :::::::::::::::::::: ::::::::::::::::::::
1657  GTGAGAGACAG.....CTAGCAAGATGAGCTGAGAGAGAGAGAGACA 1694
    :::::::::::::::::::: ::::::::::::::::::::
457  actLysLysLysL1LysL1LysL1LysL1LysL1LysL1LysL1Lys 474
    :::::::::::::::::::: ::::::::::::::::::::
1695  GCACTTGAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1744
    :::::::::::::::::::: ::::::::::::::::::::
474  spG1LysL1.....LleLysGln 479
    :::::::::::::::::::: ::::::::::::::::::::
1745  CCGAGACAGCTGCGAGCTGCTGAGAGAGAGAGAGAGAGAGCTTGAAG 1794
    :::::::::::::::::::: ::::::::::::::::::::
480  LeuG1LysPheLysLysAspValG1LysLysAspTrhLysSerG1Lys 496
    :::::::::::::::::::: ::::::::::::::::::::
1795  CCGAGTCCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
    :::::::::::::::::::: ::::::::::::::::::::
496  yLysValLysG1LysG1NhrTrhLysPheLysL1LysL1LysL1Lys 513
    :::::::::::::::::::: ::::::::::::::::::::
1845  CAGCTGGAGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888
    :::::::::::::::::::: ::::::::::::::::::::
513  hrSerAspAlaLysTrhLysL1LysL1LysL1LysL1LysL1LysL1 520
    :::::::::::::::::::: ::::::::::::::::::::
1889  CCGGAG 1938
    :::::::::::::::::::: ::::::::::::::::::::
521  .....HisAspLeuMeLLeuLysL1LysL1LysL1LysL1LysL1Lys 535
    :::::::::::::::::::: ::::::::::::::::::::
1939  GAGCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1988
    :::::::::::::::::::: ::::::::::::::::::::
535  yAlaLysTrhLysL1LysL1LysL1LysL1LysL1LysL1LysL1Lys 552
    :::::::::::::::::::: ::::::::::::::::::::
1989  CCAAGAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
    :::::::::::::::::::: ::::::::::::::::::::
552  LysG1LysL1LysL1LysL1LysL1LysL1LysL1LysL1LysL1Lys 565
    :::::::::::::::::::: ::::::::::::::::::::
2039  TGCGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
    :::::::::::::::::::: ::::::::::::::::::::
566  G1LysL1LysL1LysL1LysL1LysL1LysL1LysL1LysL1Lys 582
    :::::::::::::::::::: ::::::::::::::::::::
2089  AAG 2138
    :::::::::::::::::::: ::::::::::::::::::::
582  tLysG1LysL1LysL1LysL1LysL1LysL1LysL1LysL1LysL1Lys 597
    :::::::::::::::::::: ::::::::::::::::::::
2139  GAGCAAG 2185
    :::::::::::::::::::: ::::::::::::::::::::
597  eG1LysL1LysL1LysL1LysL1LysL1LysL1LysL1LysL1Lys 613
    :::::::::::::::::::: ::::::::::::::::::::
2186  TGAAG 2220
    :::::::::::::::::::: ::::::::::::::::::::
614  AlAg1LysPheTrhLysL1LysL1LysL1LysL1LysL1LysL1Lys 630
    :::::::::::::::::::: ::::::::::::::::::::
2221  GAGAGAG.....AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2252
    :::::::::::::::::::: ::::::::::::::::::::

```

```

630 uLysAlaLeuLysLeuLeuLeuAspValGluAArgIuTrGluLysSerLeu 647
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2233 GCGGAGCAAGAAAGACCTGGCAAGAGAGGGCTGGCGGCAAGGGCGCTGG 2302
647 sn ..... GluThrArgPro 651
2303 AGGCTGATTAATCTACGAGCTGGGAGGCGCATCCAGCTGAGACTGAGATC 2352
652 LeuLysArgGluLeuGluLysLeuLeuLysLeuLeuLeuLeuLeuVal 668
2353 CTGGCGGCGGGAGCTG..... GCAAGCGCTGGCTGCCAGACACAC 2396
668 sProGluGluAspGluLeuLysSerArgLeu ..... 679
2397 TGAAGTGAAGTGAACGCTGCAAGAGACAGCTGCTGGCTGGCTGAC 2446
680 ..... GluGlnLysSerGluLeuLeuLysArgLeuThr 691
2447 GGTATGAGATGACGCAAGAGAGGCGCACAGTATGCGCGCATCTGTCAG 2496
692 GluLeuThrSerLysAspGlnThrLeuGlnLysGluLeuLysValGly 708
2497 GAA..... CAGCTGATGACTTTGAGAGAGAGATGAGAGAGAGCGCG 2537
708 sLeuAspAsnLysLeuLeuThrGlnGlnValAspAsnLysThrGluLys 725
2538 CCAAGAGCTGCAAGAGGCAAGAGAGAGCTGCGAGGCTGAGATCCGACA 2587
725 eLysGlnValProLeuLysValSerGluLysValLysSerHisAsp 741
2588 GCGAG..... CTGCAATAGCCGCGCACACAGCAACAA..... 2619
742 ValIleValAspAspLeuAsnLysLysLeuSerAspValThrHisLys 758
2620 ..... CTAGCTGAGCTTCATCCACAGCTGGCGACAGCATCTGACAGCT 2663
758 rThGluLysLysLeuGluLeuGluLysLeu.....L 769
2664 CCAAGAGAAAGAGCTGAGCGCCAGAGAGCTTCCGACATGACCTCTGACTC 2713
769 eLysGluLysAsn...AlaSerLeuSerLysAsnValSerArgLeuGlnThr 784
2714 TGCAGAAAGATGAGCTGCGCACAGCAAGAGTGGCGCTTGAGAGAC 2763
785 ValIlePhe..... 786
2764 TTGTGCTCCCAAGGCAAGTGAAGCAAGCAAGAACGCTCCCGGAGTTAACT 2813
787 ..... IleProProGluArgHisGluLysG 795
2814 CAAGAGAGCTGCGAGGCGCAGGAGACAGAGCCGACATGAGCTGGAGAGGC 2863
795 Lu..... MetPheAlaLeuLys 800
2864 AACAGAGAGCGCAAGCTTCTGACCAACAGAGAGCGAGCTGCGATGAGAG 2913
801 SerAsnIleThrGluLeuLysGlnLeuSerGluLeu..... 813
2914 CCGAGAGGACAGACAGATGAGCAATGAGCTGGAAGAGCTCCGAGCGCCCT 2963
814 ..... AsnLysLysCysGluLysAspGlnGlu 823
2964 GATGCAAGAGCGCAGGCGACAGCAAGGAGAGAGCTGCGAGCAAGAAAGC 3013
823 ySIIleTyrSerLeuMetSerGluLysAsn..... 832
3014 AGGTGGCGCGCTGAGCCAGAGAGGGGCGCTGCCAGGCTGACCTTGCC 3063
833 ..... AspLeuLysLysThrMet 838
3064 CTGGAGAAAGCGCGCCAGAGCAAGAGCTGAGATGCGGCTGCAAGAACGCCCT 3113

```

---

```

838 tSerHisGlnTyrValProValLysThrHisGluGluLysThrAlaL 855
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3114 CAAGAGCAAGAGTGTGGAGTGGCTGGCTGACCTCCAGAAAGGCGAGGCTCATG 3163
855 euSerSerThrLeuAspLysThrAsnArgGluLeuLeuValAspValLys... 870
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3164 CCTGACGGAAAGAGAGCAAGAGACACAGAGTATGGCCAGACCTGCTGT 3213
871 ..... LysLysCysGluLysIleAsnGlnGluPheVal 881
3214 CTGAGAGCGAGCCCAATTAAGACCTGAGAGAACTTGGCAACACCTGAA 3263
881 LysIleLysAspGlnAsnGluLysLeuLysValAspAsnLeuGlnAsnThrG 898
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3264 GCAAGTGAAGGAA..... CAGCTGCTTAAGAAAGAAAGAGAGCGGAT 3307
898 IAsnGln..... 900
3308 CTGGCTGAGAGCCCAATCTGAGAGCTGCTGAGAGAGAGCAAGCAGC 3357
901 ..... ValValIleGluTyrIleSerLeu..... 908
3358 CCGAAGCTGAGAGACCTGGCGGCGAGAGCGAGAGCAAGCTGAGAACGATG 3407
909 .ArgGluHisGluGluLysMetSerLysLeuArgLysSerMetLysVal 925
3408 CCAAGAGCAAGAGAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGCTG 3457
925 aGlnAspAsnSerIleGlu..... IleuAla 934
3458 AGCGGCGCTCCGCGCTGAGAGCGGAGAGATGCTCTGAGAGCTGCAAGGC 3507
935 LysTyrLysLysSerGlnGluGluIleValThrLeuHisGluLysIleAl 951
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3508 CAGTTAGAGAGAGAGCCAGAGAGCTGAGAGCAAGTATGAGATGCTTAC 3557
951 aAlaGlnLysArgGluLeuAspAsnThrIleGlnGluLysGlyLe 964
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3558 CTGGCGCAAGAGAGTGGCTGCTGCTGCGACCAAGGTAAAGACACAC 3607
964 ..... 964
3608 GCAAGCTGAGATGAGTGAAGAGCCAGAGTGGCGCGGCGCCGCGAGAGC 3657
965 ..... LysIleuLysTyrAlaProIleIleSerLeuGlnGlu 976
3658 GCTGAGAGAGAAATAGCTCTATGACAGCAAGTGGAGAGAGAGCTGCAAT 3707
977 .CysGluArgLysPheLysAlaIleThrGlu.....LysGluLeuLysG 990
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3708 CTTGAATGCGAGGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGATTTGAAGC 3757
990 LysIleuSerGlnGlnThrGlnLysTyrAsnThrSerGluGluLysAla 1006
3758 GCGTGTGATGCGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 3807
1007 LysLysCysLysGlnGlnLysAspLysLysLysLysLysGluLysIleuThrIle 1023
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3808 GCTCTGTCAGGCGAGAGACACCAAGAGTGGCGAGAGAGCTGCAAGACC 3857
1023 uGln.....LysAspLeuLysA 1039
3858 AGCTGCTCTCTGCGAGAGAGCTGACAGAGCTCCGGAGAGAGAGAGCTGAGA 3907
1029 sPlyAsnValHisIleGlnLysMetTyrGluThrGluAlaGlnLeuSer 1045
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3908 AACACCGGCTGCTTACAGAGAACCTGCGAGAGAGAGAGTCACTTCAAGAGCT 3957
1046 ArgLysThrGluLysAsnArgLysLeuLysAspLeuGlnLysLys 1062
3958 GAGCTGCGAGAGAGAGCTGAGCAAGATTAAGAGAGCTGCGCAAGAAAGTT 4007
1062 rThrGluAlaLysGlnLysGluLysLeu...ValIleGlnLysAlaL 1078

```



alignment scores:  
 Quality: 512.00 Length: 1541  
 Ratio: 0.660 Gaps: 61  
 Percent Similarity: 50.357 Percent Identity: 20.052

alignment block:  
 US-09-750-590-2 x PCT-US93-06160-3

Align seg 1/1 to: PCT-US93-06160-3 from: 1 to: 6306

```

185 LysGIlnAsnATgThrAlaLeuMetLeuGlyCysGIuTyTcLYuLYsAs 201
      ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|
643 CAGATGAGAGCGCTGAGAAAGACAGCTGCTGATGACAGAAAGATATGAGGA 692
      ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|
201 PALValGluVal..LeuLeuLysAndLYuAlAspAryLThrLeuLeuA 217
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
693 TGAGCTGGAGCTGGAGCTTACAGTGGAGAACCGAGCTCTCAGCCAAAGAG 742
      ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|
217 sPALLeuLcLYuAlAspSerSetTyTyrAlaLarTgLeuGlyAspAInLc 233
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
743 ATGCA.....CAGATGAGCATATGTCAGACCGCAT 774
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
234 AspILeuThrLeuLeuLYuThrAlaSerGluAsnSer..... 246
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
775 GACCGCTAGCGCTGTAATGAGAACAGCGCGCCAGCCAGCCAGAGCC 824
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
247 ..AsnLYuGIuTyTgLYuLeuTrLYuSGLYProSerLeuGlnGlnAry 262
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
825 CAGAGAGCTGAGAGCTGCGTGACAAAGATGAGAGCCTTACCATGCGGC 874
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
263 .....AsnLeuSerGlnMetLeuAspGluValAsnThrLYuSerAsn 276
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
875 TGCATGAACCGCTGAAAGCGATGCGAGAGC...CTGGAGACAGAGAAAGAC 921
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
277 GlnATgGlnLysGlnAsnILeGlnAspLeuLcLYuLysAsnGluAspLe 293
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
922 CAGATGCAT...CGCAAAATCAACCGAGCTTCGAGAGAGATGAGACCT 968
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
293 LYuSGLuTyGLeuAryLYuSILeGlnGlnGluGlnAryLLeuLeuAspL 310
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
969 TTCCTTAAAGTTCGGAGATTTGCCAGATCATCTGACAGAGCTAGAGATG 1018
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
310 yValALAsnGLYLeu..... 314
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1019 CCGTCATATGAGCTGACGAGAGACAGACAAAGCGCATCTGACAGATGCGCTA 1068
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
315 .....GlnLeuGlnLeuAsnGlnGluValALAspAspLeuGln 329
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1069 GAGAGACAGCGCGAGCTGAGAGAGAGAGAGCTGACAGAGAC...CTGCA 1112
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
329 uSerGluLYs.....GluLYuLeuLYuSerLeuLeuAlaLYuGluLYu 344
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1113 GCAAGAAAGATGCTTGAGAGAGAAAGCAAAATCTCTCAGGGAAACTTT 1162
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
344 ySGlnILsGlnLysLeuAryThrILeGln..... 354
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1163 CAGAGCTGAGAAAGACATCTGTCGCAAGTGTGAGATAACCCAGCCAGAG 1212
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
354 ..... 354
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1213 AAGGGAGAGCTGAGTGGATGCTTGACATGAGAAACCTGAAAGAGGA 1262
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
355 .....AlaLeuLYuSerATrPheLYuT 362
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1263 GCGACAGCACTCTTGCTGCAAAACAGACAGAGCTCCAGAGCGAGTGAAGA 1312
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
362 yTrPheGluSerAspHis.....LeuGlySerGly 371
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1313 TGCTGGAGAGTGAAGCGCCAGAGAAACCAAGCTGCTTCTGAGCGGG 1362
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
372 SerILAspAryLYuSGLuAspMetLeuLeuLYuGln..... 383
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

1363 GCGCACTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1412
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
383 ..... 383
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1413 GCAAGCTCATCTGCAAACTGACAGCGCCAGAGAGAGAGAGAGAGAGAG 1462
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
384 .....GluGlnMetLYuMetThrAsp 390
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1463 CCGCCAGAGCTATGGGCGCGTTAGCTAGCTGCGGCTGAGCTGAGCC 1512
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
391 SerGlnCyTrPheSerThrGlyMetProValLysMetGlnSerATySerLc 407
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1513 TCTGAGCTGACAGCACTCAATGCCAGATCTGACAGAGAGAGAGAGAG 1562
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
407 LLeuATrProLYuGlnLeuLYuAlaLeuTrAspAsnGlnLAserTySerGln 424
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1563 GAGTGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1612
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
424 sNGILcLYuLeuLYuSGLcLYuGlnLcLYuAlaMetLYuThrPheCysAspSer 440
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1613 CC.....CTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
441 ALALysGlnAryLarGlyLYuSGLAsnAsnLYuAlnLysLYuAlAL 457
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1657 GTGGAGACAG.....CTAACAGATAGCTGAGAACAGAGAGAGACA 1694
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
457 sGILcLYuSALeuAlaLeuGlnLYuSGLuTyGlnAryLYuSGLuAspSerA 474
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1695 GAGTTCAG 1744
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
474 sPGLcLYuLc.....LLeuGln 479
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1745 CCGAGAGAGCTGCGACAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1794
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
480 LeuGlnAspALaLeuLYuAspValGlnLYuATrMetTyTcLYuSerGln 496
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1795 CGGAGTGCAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
496 LYuVALLYuSGLMetGlnThrILAspLeuLYuAlaLYuSGLILsLeuTr 513
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1845 CAGAGCTGAGATTTGCGACAGAGCAACTTCAGAGGTGAGTAATGAA.....G 1888
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
513 TrSerAspALaLarThrGlyAsn..... 520
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1889 CGCGAGAGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
521 .....HisATrLeuMetGlnLYuLeuLYuAspTrGlnLYuAspMetLYu 535
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1939 GAGCTGACCGAGAGAGTGGAGAGATCTGACAGAGCTGTGTGAGAGAGCGCC 1988
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
535 sVALyTyTgLYuLYuAspATrAlaLYuValGlyLYuValTrAspAsnGln 552
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1989 CAG 2038
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2039 TCGGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
552 LcLYuGlnALAsnGLcLYuMetLeuValGILcLYuThrLYuATr.....Asp 565
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2093 TCGGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
566 GILcLYuLYuMetGILcLYuLAsnLYuATrLYuGlnLYuGlnLYuSerLc 582
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2089 AAG 2138
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
582 TCySGL.....LeuGlnAryLYuAspLeuSerATrAlaLYuLAsnLYuSer 597
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2139 CAGCAAG 2185
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
597 eTcLYuGILcLYuLeuLYuAspLeuSerATrAlaLYuLAsnLYuSerLc 613
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2186 TGAG 2220
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
614 ALcGLcLYuThrGILcLYuAsnMetLYuSerLeuLeuSerATrGlnLYuAsnGln 630
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2221 CGAGAG.....AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2252
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

[illegible]







```

2033 ..GATGACGACATACGAAAAATTAATCTAGACGACATGATGATTAAC 2080
|||||
630 Glu.....LysAlaLysLysLeuIleAspValGluAsp 640
TCCCTGTGGGATATTTTCCCAACAAAACGCTTGAGACCTGGCTGA 2130
640 9GLuTyrGluAspSerLeuAsnGluThrArgProLeuLysArgGluLeuG 657
|||||
2131 TAGTAATCAAAAAGAAATTATATCAGCCAGGAC.....ACATTTG 2171
657 LysAlaLeuLysAlaLysLeuAlaGlnIleValLysProGluGluIle... 672
|||||
2172 CCAAATTGACAGACAGACACTAGCTTCATCTGACGACGAAATAAATCATTA 2221
673 ..GluGlnIleLysSerLeuAspGluGlnGluLysSerGlyLeuGluGly 688
|||||
2222 AATATGACATCAAAAAGACAGACAGACAGCTTGTGCTGACAGACGA 2271
688 sArgIle.....ThrGluLeuThrSerLysAsnGlnIle... 700
|||||
2272 GCGTGTGATGTTTGTTGGTACCCAGCATTTTGCAATATATTAGACAGCG 2331
700 euGlnLysGlnIleGluLysValLysLeuAspAsnLysLeu..... 714
|||||
2332 TTAAGAGGAATTTAAATCAATCAACAAAACGACGACCTGGCTGGA 2371
715 .....ThrGlnIleValAsnAsnLeuThrThrGlnLeuLys 726
|||||
2372 GCCACACGACGATTACTCCGACGTTCACTACTACGTACACAGCAAAACGA 2421
726 sAsn.....ValProLeuLysValSerGluGluLysSerThr 740
|||||
2422 GTCATGTGCCCGCTTTGTCAAGAGGTTTTCAGACAGAGCTGGATAC 2471
740 LAspValIleValLAspAsnLysLysLysLeuSerAspAlaThrHis 756
|||||
2472 AAGAAGTCATC...AGTATTTCCGACGTCAAACTG..... 2503
757 LysTyrThrGluLysLysLeuGlnMetGluLysLeuMetGluAsnAl 773
|||||
2504 .....GNCCTTCCTCCAGATGACGATGACGATCAACAGCATC 2538
773 aSerLeuSerLysAsnValSerArgLeuGluThrValPheIleProG 790
|||||
2539 AGACCTAAAAAAAAG.....G 2555
790 LAspGlnIleLysGlnMetMetAlaLeu.....LysSerAsnIle 803
|||||
2556 AAGAGCGCGGATGATTAATCTGGACCTGGCCCATACGCAAGCAATA 2605
804 ThrGluLeuLys...LysGlnLeuSerGluLeuAsnLysLysGlySerGly 819
|||||
2606 ATTGATTGTAAGGAGAGCAATACACGAAATTAGCAAAACAACTGCA 2655
819 uAspGlnGluLysIleTyrSerLeuMetSerGluAsnAsnAspLeuL 836
|||||
2656 TGTCAATATAGACATACAGCGGCTTA.....AAGACGACATGTAAG 2696
836 yThrMetSerHisGlnIleValProValLysThrHisGlnIleLys 852
|||||
2697 AACAAAGAACACCTCTGGTACATATATCCCTGACAAAGAAAGTCCAA 2746
853 ThrAlaLeuSer.....SerThrLeuAspLysThrAsnGluLeuVal 867
|||||
2747 GTATGCTGTACAGATGTTACATATTATAGAGGCTTCAGATCAACTTA 2796
867 LAspValLysLysGlySerGlu.....A 875
|||||
2797 AGATGTTGAAAAGAAATTCGACACAGCAAGCAAGCTACGTCACAGATAG 2846
875 sPLeuAsnGlnGluPheValLysIleLysAspIleAsnGluIleLys 891
|||||
2847 ACTTAGCTGAACCTGCCACACAGTACACGAGGAAACAGCAAGAACG 2896
892 ArgAsnIlePheGluAspThrThrGlnAsnGlnValLysAlaGluTyrIleSerIle 908
|||||
2897 CACAAGTTGACACAGGTTTGTAGTAAATTAATTAATTAATTCATAGCTAT 2946
908 uArgGlnIleGlnIleLysMetSerGlyLeuLysGlySerMetLysLys 925
|||||
2947 ACAGACACAGCAGCAGACATCAATCTAAAAAGTCAACAAATATACC 2996
925 aGlnAspAsnSerAlaGluIle.....LeuAlaLysTyrLysLys 938
|||||
2997 TAAATCTGAGAACTTCATATATCCATTAATTTGACAACTGTCCACAA 3046
939 serGlnGlnGluIleValThrLeuIleAsnGlnIleAlaIleAlaGlnLys 955
|||||
3047 CTGAGAGCAGCAGCAGCTGTGAATTTACTACCTACAGTTGAGTGTGAC 3096
955 gGluLeuAspThrIleGlnIleLysIleLysLeuLysTyrAlaProIle 972
|||||
3097 AGATTAAGATCCCTAAAGG.....CAGTATCA 3125
972 LserLeuGlnGluLysGlnLysValLysLysAlaThrGlnLysLeu 988
|||||
3126 GCCCTTGGCAACACATTTGCAAAAGTCCACGACAAAGAAACAAATTA 3175
989 LysGlnLeuLeuSerGlnIleThrGlnLysTyrAsnThrSerGluGlu 1005
|||||
3176 .....ATCACAAAGAAATTAATACAGCAACAAATC 3204
1005 uAlaLysLysCysLysGlnLeuAsnAspLysLeuLysGluIleLeu 1022
|||||
3205 ACCACAG.....GATTAACG..... 3220
1022 hLeuGlnLysAspLeuLysAsp.....LysAsnValHis 1033
|||||
3221 .....AGATATTAAGAGAGAGAGCTTAAAGATTAATCTGGCTATAG 3262
1034 .....IleGluAsnSerTyrGlnThrGlnLysGlnAlaLeuSerArgLys 1048
|||||
3263 AAGACATTGAGAAATTATATCAAGATGCAAGAAAGCAGCTAAGAAAGCA 3312
1048 rGluGlnLeuAsnArgGlnLeuLysAspLeuLeuGlnLysTyrThrGlnAla 1065
|||||
3313 AAGAGAA.....ACGAACTTAATTAAGTAAATACCTCAAGTGAAT 3356
1065 LAlaLysLysGlnLysGlnLysValGlnLysLysAlaLysGlnThrSer 1081
|||||
3357 GCGAGAAACCAACAAAGAAATTAATAGAT..... 3388
1082 GluIleLeuAlaIleGlnThrLeuLeuGlnLysGlnIleValProLeu 1098
|||||
3389 .....ATGACACTATGACGACAAAGATTTGATGACAA 3420
1098 uGlnValAlaLysSer.....LeuLysLysSerLeuSerGlyThrIleGlu 1113
|||||
3421 GAAGATACAGAAAGTGCCTCAAGATACCTT.....ACTTTAGCA 3464
1113 hLeuLysGlnGlnLys.....ThrLysGlnArgCys 1124
|||||
3465 AAGCAATGAGCAACTAAAGCACTTGACAGCAAGCAAAACACAACTTG 3514
1125 TyrGlnLysGlnGlnIleThrValThrGlnLeuArg.....GlnMet 1138
|||||
3515 AAGCAATGAGTCAATATTCAGGTTTGCAATTAAGCAAGTCAACATCA 3564
1138 LLeuGlnLysn.....GlnLysAsnSerSerValProLeu 1150
|||||
3565 GTTGAGAGAGACATACAGCAATATATAAAGAAATCATATATTTGGCATTAG 3614
1150 LAlaGlnLysLeuGlnValLysGlnAlaPheGlnLysGlnValGlyIleIle 1166
|||||
3615 GCGCA.....CAAGAGGTTTATACAGAGAAATTTATTCATTTT 3652

```

```

1 COMPUTER READABLE FORM:
2 MEDIUM TYPE: Floppy disk
3 COMPUTER: IBM PC compatible
4 OPERATING SYSTEM: PC-DOS/MS-DOS
5 SOFTWARE: Patent In Release #1.0, Version #1.25
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: 7/2000/08/687, 080
8 CLASSIFICATION: 435
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 08/592,126
11 FILING DATE: 26-JAN-1996
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Shultz, Charles K.
14 REGISTRATION NUMBER: 38,615
15 REFERENCE/DOCKET NUMBER: 4600-0111.30
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: (415) 324-0880
18 TELEFAX: (415) 324-0960
19 INFORMATION FOR SEQ ID NO: 44:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 3893 base pairs
22 TYPE: CDS
23 STRANDEDNESS: double
24 TOPOLOGY: unknown
25 MOLECULE TYPE: cDNA to mRNA
26 HYPOTHEetical: NO
27 ANTI-SENSE: NO
28 ORIGINAL SOURCE:
29 INDIVIDUAL ISOLATE: RAD50 cDNA SEQUENCE, CDS: 389 TO
30 INDIVIDUAL ISOLATE: 4324
31 IS-08-687-080-44

```

```

alignment_scores:
  Quality: 474.00      Length: 1421
  Ratio: 0.647         Gaps: 70
  Percent Similarity: 51.583  Percent Identity: 22.027

```

US-09-750-590-2 x US-08-687-080-44 ..  
Align seg 1/1 to: US-08-687-080-44 from: 1 to: 5893

[illegible]

```
seq_documentation_block:
  sequence_id: 05912_6/products/2/line/58.COMB-seq-US-08-687-080-44
  accession: 5965427
  general_information:
    applicant: Gregory Dolgano v
    title_of_invention: Human RAD50 Gene and Methods of Use Thereof
    number_of_sequences: 175
    correspondence_address:
      address: Gehlring & Associates
      street: 30 Woodbridge Avenue, Suite 250
      city: Palo Alto
      state: CA
      country: USA
      zip: 94306
```



[illegible]



```

289 ..... ||||| ||| ||||| .....
    .....CAGCGTATACATCTTAATTTTATTATACAGCAACAACTACACAGCTGC 335
369 yser1yserrhpharfglysluaspmetleu.....leu1 382
    :||| :||| :||| :||| :|||
336 TACACAGATTAACGTAATTAAGACAACTTAATTAATTCCTACGACA 385
382 yslnglucimetyrmettharp.....serlucytrhsettrh 396
    :||| :||| :||| :||| :|||
386 AAAAGCAACGCTGACACACGCAAGATTAACCATTAAGAAACGATT 435
397 glymetrprovalrh.....metlinserrgse 406
    :||| :||| :||| :||| :|||
436 AAGAGAGGCTGCATTAATTTACACACTGCAACCAAGAAAGTATTC 485
406 rhg.....leuarprouleu1leu1leu1leu1leu1leu1leu1 421
486 TATGAGCAAAATCGCAACAAATTCTAACGCTAAGAGCTTTCCTAATT 535
421 yrrsergluamglu1leu1leu1leu1leu1leu1leu1leu1leu1 433
    :||| :||| :||| :||| :|||
356 TCAACATCTCGAAACTATGCTAAAGCAATTAATTAATTAATCTAAAC 585
434 metartrhrhescyarserrseralalygl1nararfglyleu1leu1 450
586 CGATTAGCTTTAGCGACATCGAAA.....ACAAAGCTTTATTAACAG 632
450 ngluleu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 461
    :||| :||| :||| :||| :|||
633 TCAACTATTGCAAGAGCGCGACGATCATAGAAAGCGCAAGCGACATA 682
462 .....leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 473
683 AGACAACTATGCACTACCTTAA..... 708
474 Aepglu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 490
    :||| :||| :||| :||| :|||
709 GATCGACAAATTAAGTACGAGAGTACGACAAATTAAGCAAACTTAATCTGTA 758
480 gmetrtylucserglucyluvalalylngl1meto1nth1leu1leu1 507
759 AATTGTGATTAAGAAAGTGAATTAATTAATTAATTAATTAATTAATTA 808
507 lalculysgl1u1h1leu1th1serarplalalr1cyluan1h1n1g1leu1 523
    :||| :||| :||| :||| :|||
809 AATTTAAGCAAAATCGCAACGCAAACTCACT.....AATTGCAAAAGA 852
524 metoluc1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 540
853 GCAAGAGTAAAGTACCTTCTGCAAAATTTTAACTTAACTTAACTTAACT 902
540 y1aser1alagl1uval1gl1leu1leu1leu1leu1leu1leu1leu1 557
    :||| :||| :||| :||| :|||
903 CAAITTCGATTAAGCTTAAC.....ATTAACTAAATTAAGAAATTTAA 946
557 e1leu1val1gl1u1leu1leu1leu1leu1leu1leu1leu1leu1 565
    :||| :||| :||| :||| :|||
947 CAATTTCTGCAACATTAATTAAGATTAAGAGACGCAAAACCAAGCT 996
565 ..... 565
997 ACTCAAAAGTAGACACTATTCATTAACATGCAACCACTGACTGATTAAT 1046
566 .....gluc1lyleu1leu1leu1leu1leu1leu1leu1leu1leu1 580
    :||| :||| :||| :||| :|||
1047 TTTCAGAGAACTCGAAATTTTGAAGTGGAACTGAATTCGCAAAAAA 1096
580 e1u1ser1hpharfglyleu1leu1leu1leu1leu1leu1leu1leu1 596
    :||| :||| :||| :||| :|||
1097 TTAAATTAAGCAACGCAATTAAGCAAAATTAAGCTGACCTTCTGAA 1146
597 metoluc1gl1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1 613
    :||| :||| :||| :||| :|||
1147 ATTCAAAACAGCTGAAAGGTTTCGAAAGAAATTAATTAATTAAGCA 1196
613 oalagl1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 630
    :||| :||| :||| :||| :|||
1197 GCCCAATTAATTAAGCAACGCAAGAGTGAATTAATTAATTAATTAATTA 1246
630 l1u1val1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 646
    :||| :||| :||| :||| :|||
1247 GAAATTAATTAAGCAACGCAAGAGTGAATTAATTAATTAATTAATTA 1278
647 AANG1uthratarprouleu1leu1leu1leu1leu1leu1leu1leu1 1280
    :||| :||| :||| :||| :|||
1279 GAAAGCTGAAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
663 u1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 680
    :||| :||| :||| :||| :|||
1321 .....CAGAGAGCAAAAGAAATAT 1339
680 luc1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 696
    :||| :||| :||| :||| :|||
1340 ACAGTAAACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1386
697 AANG1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 713
    :||| :||| :||| :||| :|||
1387 AAT.....ACTGCAAGAAAGCTTAAAGCAAGCAAGCAAGCAAGCAAG 1433
713 u1eou1th1ng1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1 730
1434 TATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1474
    :||| :||| :||| :||| :|||
730 eul1u1val1sergl1u1leu1leu1leu1leu1leu1leu1leu1leu1 745
1475 TATCTGTTACGAAAGAAATTAAGCAAGCAAGCAAGCAAGCAAGCAAG 1524
746 Aap1eul1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 762
1525 AATTGCAAGGCACTTAATTAATTAATTAATTAATTAATTAATTAATTA 1574
762 sleng1umetol1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1 779
    :||| :||| :||| :||| :|||
1575 TTTCGAAAGCTTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1624
779 al1ser1ar1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 795
    :||| :||| :||| :||| :|||
1625 TACCAAA.....AAATCA 1638
796 metleu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 806
    :||| :||| :||| :||| :|||
1639 ATTAAATTAAGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1688
807 .....lysluc1leu1leu1leu1leu1leu1leu1leu1leu1leu1 818
    :||| :||| :||| :||| :|||
1689 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1738
818 lyl1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 834
    :||| :||| :||| :||| :|||
1739 GAGAA.....AAGCAATTT 1758
835 lyl1u1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 851
    :||| :||| :||| :||| :|||
1759 GAGAGCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1790
851 e1u1st1n1al1eul1ser1leu1leu1leu1leu1leu1leu1leu1leu1 860
    :||| :||| :||| :||| :|||
1791 AAGAAATGCTTCGCAACCTTCAGCTTAATTAATTAATTAATTAATTAAT 1840
861 .....lyst1h1nar1ar1gl1u1leu1leu1leu1leu1leu1leu1leu1 875
    :||| :||| :||| :||| :|||
1841 GCGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1890
876 l1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1leu1 889
    :||| :||| :||| :||| :|||
1891 GCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1940

```



890 .....LeuLYsARgAsnLeuGluAsnThrGlnAsnGlnValIleAlaAG 904  
 1941 ATTAAATATGAAAAGAGCTTAGATAGATTATATGATGATTAATAAG 1990  
 904 LuYrLIeSerLeuArgLIuHISgluLIuLYsMetSerLIuLeuArgLYs 920  
 1991 AACACATACAGCTACATGAGACACGCTAGCTATTAACACACTGCTGA 2040  
 921 SerLeuLYsLYsValGlnAsP.....AsnSerLIuLIuLeuLIuValY 935  
 2041 AATATTAAAGCTAGTGTGATGAAAACACTTGTACATATTATTAAAAA 2090  
 935 sTYrLYsLYsSerGlnGluLIuLIuValTrpTrpLeuHISgluLIuLeaIa 952  
 2091 AGCAGCGGACAAATATGAGAAATACAAATACCACTCATCCCTTA 2140  
 952 LAGLIuLYsARgLIu.....LeuAsnTrpLIeGlnLYs 963  
 2141 AAAATTAAGCAAAACATTTTAGACACGTATACCTCTCTGCAAAATAT 2190  
 964 LIeLYsLeuLYsTYrAlaProLIeLIeSerLeuGlnLIuLYs.....GI 978  
 2191 GTAGCGCATTTAAATATACCCCTGACTGTGCTACACAAATTTTAGAGA 2240  
 978 uArgLYsPheLYsAlaThrGluLYsGLuLeuLYsGln..... 991  
 2241 TCCAAATTTGAA.....ACTCCAGCAATTTAAATTTGAGCTGAAAAATA 2287  
 992 .....LeuSerGlnLIuThrGlnLYsTYrAsnTrpSerLIuGlnLIu 1005  
 2288 ATGTACTACTAGAACACGAAATATGTCAAGAAATACCAACGATGTGAT 2337  
 1006 AlalYsLYsCYsLIuGlnLIuAsnAspLYsLeuLYsLIuLIeLeuTh 1022  
 2338 GTTCATTAATAATATACAGGCTTACAGAGCTGACGAGCAATCTTCG 2387  
 1022 rLeuGlnLYsAspLeuLYsAspLYsAsnValHISLIeLIuAsnSerTYrG 1039  
 2388 CCACCTACAGCGAAATA.....G 2404  
 1039 LuThrGluArgAlaLeuSerArgLYsThrGlnLIuLeuAsnArgLIu... 1054  
 2405 ATACAAACAAACAAACAGCTATCAATATATATACAAAGGAAACCAATA 2454  
 1055 .....LeuLYsAspLeuGlnLIuLYsTYrThrGlnAlaLYs..... 1066  
 2455 TATCTTAAAGTGTGCTTAATATCAATATCAAAAAATTAATACTCTAT 2504  
 1067 LYsGluLYsGLuLYsLeuValGlnLIuAsnAlaLYsGlnTrpSerGluI 1083  
 2505 AAAAAGTAAAGGAAAGAGCTGTTCCAGTCAAAATAGCTATTC.... 2550  
 1083 LeuLeuAlaLIeGlnThrLeuLeuGlnLIuLYsGlnHISValProLeuGln 1099  
 2551 .....ANGAAACATAGCGAGTTAGCAAA 2574  
 1100 ValGluSerLeuLYsSerLeuSerGlyThrLIeGluThrLeuLYsGL 1116  
 2575 ATTTACATGACCGATTAATAATACATACATACATAGGTTAGAGAAACA 2624  
 1116 u.....GluLeuLYsThrLYsGlnArgCYsTYrGluLYsGlnGlnLIuThYr 1132  
 2625 AACGTAAATTCAAAATCTACCAATCTTCTACTCAAAA...AGACATA 2671  
 1132 aLTrpGlnLeuArgGlnMetLeuGlnLYsGlnLYsSerSerValPro 1148  
 2672 ACAGCAATAGCGATTCGAACTGTGAAAAATTAATAACGATTTGAAAGT 2721  
 1149 LeuAlaGluHISgluGlnValLYsGluAlaPheGluLYsGlnValGlyLI 1165  
 2722 TTGAAAAATTCATTA.....AAACACTGAAAGGAAACTTAATTC 2762

1165 eLIeLYsAlaSerLeuArgLIuLYsGluLIuGlnLIuSerGlnAsnLYsThrG 1182  
 2763 TCTTAAAGCAAGC...TCCGCAATCATGAAACCTGTACAAAGTAAAGTCG 2809  
 1182 LuGIuValSerLYsLeuGlnSerGluLIeGln..... 1192  
 2810 AACCCATTAATCCCTGCCCTATCCCAATACAAAGAAAGAAACGCACTA 2859  
 1193 .....AsnThr.....LYsGlnAlaLeuLYsLYsLeuGlnThrAr 1204  
 2860 CATAGCTTAATATCCCGCTGATGAGATTTATTAATAAGAAAGCAAGCATG 2809  
 1204 gLIuValValAspLeuSerLYsTYrLYsAlaThrLYsSerAspLeuGluTr 1221  
 2910 CGAAAGTA.....TCTAGGTCAAAACGTATTAAGATACCTTGCCA 2850  
 1221 ThrGlnLIeSerAspLeuAsnGln..... 1228  
 2951 AAGAAATAGATGATACACCGATTAACCAACTATACAGCAAAATGCT 3000  
 1229 .....LYsLeuAlaAsnLeuAsnArgLYsTYrGluLIuValLYsGlnLI 1243  
 3001 AAACACTACTGCGCTATTTATTAATAAAATTAATGAAACACAGCGAGA 3050  
 1243 uValLeuHISAlaLYsLYsLYsGluLeuSerAlaLYsAspGluLYsGlu 1260  
 3051 TGTCTTCACTTAATGCG..... 3069  
 1260 euLeuHISPheSerLIeGlnGlnLIuLIeLYsAspGlnGlnLIuArgLYs 1276  
 3070 .....CATTTCAATACAAACAGGTAGATATACAGCGCACTAATTT 3114  
 1277 AspLYsSerLeuThrThrLIeThrGluLeuGlnArgLIeGlnLIuSe 1293  
 3115 GATTAATCAATATACATGTCGCAAGAGTATACACAGCTACTACGTC 3164  
 1293 F..... 1293  
 3165 AAAACATTAATAGTAAACCTAAAGGTATATTATGAAAGTTAGCAAA 3214  
 1294 .....AlaLYsGlnLIeGlnAlaLYsAspAsnLYsLIeThrGluLeu 1307  
 3215 ACAGCTGAATATCACTATCAAGACAGTGCAGAAAGAAATTCAGCTTC 3264  
 1308 LeuAsnAspValGluArgLYsGlnAlaLeuAsnGlnLYsLeuSerGlnLIe 1324  
 3265 TATACGAATTTAAAAATTAATAAAACATCATTTAAACGAAATTTATCAAC 3314  
 1324 u..... 1324  
 3315 ATCAAAATGAATTAAATTGCAAGAAATGAATCAAAATGCTATTAATACA 3364  
 1325 .....ThrLYsLIeSerLYs 1329  
 3365 TCGATGTATCTAAATATTTATACACGTATTAACGCTCAAAATGCAAT 3414  
 1330 SerProSerLYsArgGlnSerGlnLIuLeuLYsSerLeuGlnGlnLIuA 1346  
 3415 ATAGTAACATATCAACATAC...ATAAACAATGTTAAAGCAAAAT 3458  
 1346 LArgSerLeuGlnGlnLIuLeuAlaAspLIaAsP..... 1357  
 3459 AAAAGAAAGCTACAGAAATTAATTAAGCGCTGACGACGCTTACTTACATAG 3508  
 1358 .....ArgGlnHISgluValLIeAlaLIeLYsArgThrHISLeu 1371  
 3509 ACTGCATTAATAAAGTTTAACGAAATATATACGTATTAATTAAGCTATATA 3558  
 seq.name: /cgn2.6/prodata/2/ine/08.COMB.seq.us-09-104-324B-1  
 seq.documentation.block:  
 ; Sequence 1; Application US/09104324B  
 ; Patent No. 6232460



```

608 LeuAlaLeuSerIleProIleGluIysPheGluIuAsnMetLysSerLeuLeu 624
    ||| ||||| :||| ||||| :||| ||||| :||| ||||| :||| |||||
1085 TTACAAAGAGGTGAGTACTCAAAAG.....GCCTT 1116

624 userAsnGluLeuAsnGluIuIysAlaIysLysLeuIleAspValGluArg 641
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1117 AGAGAAAGATTACAGATTCAGAACAAACAAATTTGCTGAGCTACTGAAAG 1166
    || ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1167 AAAAGAAAGCTAAATGACAAAGAAATCTAAAT..... 1195

658 AsnLeuIysAlaIysLeuAlaGlnIleValLysProGluGluIleGluGlu 674
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1196 .....AAGCTAAAGAGCTGCAATTCCTTTGCTGACTGACTGAAAC 1239

674 nleuIysSerArgLeuGluGluIuIysSerGluIuLeuGluIysArgIle 691
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1240 TACTGTGCACTTCGAAAGATTA..... 1264

691 hrcIleuThrSerLysAsnGlnIhrIleuGluIuIleGluIysVal 707
    ||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1265 .....TTGAGAACAGAACAGAACAGAACAGTAAAAAATGAAAGCAATTCG 1309

708 CysLeuAspAsnIuIysLeuIleThrGluGlnAlaIuAsnAsnLeuThrIle 1347
    ||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1310 .....AAATACTTACATGAGAGCTTAAAGAAATCAAGATGA 1347

724 uMetLysAsnValProLeuIysValSerGluMetLysLys 738
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1348 GCTG.....GAGAGATCACTAAAGCTTATGCA 1373

739 ..SerHisAspValIleValAspAspLeuAsnIuIysLysSerAspVal 754
    ||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1374 ATACAAAGAGATTCAGACTGAGATTCAGAAATTCAGAAAGCTCTGGA 1417

755 ThrHisLysTyrThrIuIysLysLeuGluMetGluIuIysLeuMetGlu 771
    || ||| ||||| || ||||| || ||||| || ||||| || |||||
1418 .....GAAAGCAAGAACCTTTATATGCA 1440

771 uAsnAlaSerLeuSerLysAsnValSerArgLeuIuThrValPheIle 788
    ||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1441 AATTAACAAATTCAGATTCCTGAAATTCAGAAATTCAGAAAGCA..... 1483

788 GTPGGLIATGHSGLIuysGluMetMetAlaLeu.....Lys 800
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1484 .....GACAGAACTAATTCGCTCTCCAAAGCCAGAG 1519

801 SerAsnIleThrGluLeuIysLysGlnLeuSerGluLeuAsnIuIysLys 817
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1520 AAGAGAGATCATGCTTGGAAATACAGATTCAGCCATTCAC..... 1561

817 sGlyuAspGlnGluIuIysIleTyrSerLeuMetSerLysAsnAspMet 834
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1562 .....ACAGATGACAGATTAATTC.....AAGAGCTTAAAGATC 1598

834 euIuIysThrMetSerHisGlnTyrValProValIuThrHisGlu 850
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1599 TAAAGACTGAGCTGAA.....AAGAGAG 1624

861 IleLys..ThrAlaLeuSerSerThrLeuAspLys.....ThrAs 1663
    ||||| ||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
1625 CTTAAGAAATTCGATTACTTCAGACTGCAACAGCTTCCTCACTGAAAG 1674

863 nArgGluLeu.....ValAspValLysLysAla 1672
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1675 CAAGAGCTTCAGACAGAAACAGTATGATGACCTCAACAGCAAGCAAGTC 1724

872 yscGlyAspIleLeuGlnIuIysIleValLysIleLysAspGluAsnGlu 888
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1725 AGCAGAGATTAATTAAT.....ACAAAGAGAGAGAGAGAGAG 1765

889 IleLeuIysArg.....AsnLeuGluAsnThrGlnAsnGlnValLysAla 903
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

1766 AGTTTGAAACAAATTCAGAAATTCCTGAAAGAGAGAGAGAGCAATTAAGAA 1815
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
903 acIuTyrIleSerLeuArgGluHisGluIuIysMetSerGlyLeuArgL 920
    ||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1816 TCA.....CTCAATATTCAGAG 1835

920 ySerMetCysValGlnAspAsnSerAlaGluIuIleuAlaIysTyr 936
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1836 AAGAGCTTAACAGAAAGAGAT.....GAGCTTAATTCGAAATTCG 1876

937 LysIysSerGlnGluIuIleValThrIuHisGluGluIuAlaIle 953
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1877 GACAGAGCTGAAAGAAATTCATTAATTAAGCAACAGCTGAAAT... 1924

953 nIysArgGluLeuAspThrIleGlnGluCysIleLysLeuIysIleAsp 970
    || ||||| || ||||| || ||||| || ||||| || |||||
1925 .....AAAGAAAGAT.... 1936

970 rolIleIleSerLeuGluCysGluArgLysPheLysAlaThrGluLys 986
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1937 .....ATTGAAAGCTTCAGACAGAGAAAGAGAGAGAGAGAG 1975

987 Glu.....LeuIysGluIuIleuSerGlnIhrGlnIuIysTyrAs 1000
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1976 AAGGTCAGACAGAAAGCAAGCAAGCTGATTCATTCAGATTAAGCTCA 2025

1000 rThrSerGluGluIuAlaLysLysCysGluGlnIuAsnAspLysLeuL 1017
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2026 TAAATTCAGCTTACACAGAAAGCTGCAAGAGAGAGAGAGAGAGAGAG 2075

1017 yLysGluIleLeuThrLeuGlnIuAspLeuLysAspValHis 1033
    || ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2076 CAGAC.....ACATTCAGAAAGAAATTCAGCAAGAAAGATATCA 2116

1034 IleGluAsnSerTyrGluThrIuArgIuAlaLeuSerArgLysThrGlu 1050
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2117 GAGCAAGAT..... 2125

1050 uLeuAsnArgGluLeuLysAspLeuGlnIuIysTyrThrGluAlaLys 1066
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2126 .....CTTTTGAGAGAGCTTCAGAAAGCAAGAG 2153

1067 .....LysGluIysGluLysLeuValGluIuAsnAlaLysGlnThr 1080
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2154 TAAATTCAGTATGACAGCTTAAATTCAGAAAGAGAGAGAGAGAGAGAG 2203

1081 SerGluIleLeuAlaGlnIhrLeuLeuIuIysGlnIleValProLe 1097
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2204 CAGCAATTAATTCAGTAAATTCAGTAAATTCAGTAAATTCAGTAAATTC 2245

1097 uGluGlnValGluSerLeuLysSerLeuSerGluThrIleGluThrL 1114
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2246 .....AAGCAGCAATATGATTAAGAGAGAGAGAGAGAGAGAGAG 2276

1114 euIysGluGluIuLeuIysThrLysGlnArgSerGlyLeuGlnGln 1130
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2277 GAGACTGAGATTA.....GAGCTTAAATTCAGCAAGAGAGAGAG 2314

1131 ThrValThrGlnLeuArgGlnMetLeuLeuGluAsnGlnIuAsnSerSerVa 1147
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2315 GAGAGCTATCTATGAGAGATCTTGGAGAGAGAGAGAGAGAGAGAGAG 2362

1147 IProLeuAlaGlnIuHisGlnIuValIysGluAlaPheGluIysGluValG 1164
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2363 .....AAGCTGATCTTTCGCTTAAAGAGAGAGAGAGAGAGAGAGAG 2404

1164 IyIleIleuAlaSerLeuArgLysGluIuGluSerGluIuAsnLys 1180
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2405 .....AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2434

1181 ThrGlnGluValSerLysLeuGlnSerGluIleGlnAsnThrLysGlnAla 1197
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```
2435 GCAAGAGAAACACGCTACTCTTAAGAAAAAGACAGAAACAG 2484
1197 aLeuLysLeuGluThrArgGluValValAspLeuSerLeuTyrLysA 1214
      ||||| |||||
2485 AACATTTTATTGGAAACCCCTGAAATTTATGGCAATTGGATCTAAG 2534
      |||||
1214 IaThrLysSerAspLeuGluThrGlnIleSerAspLeuAsnGluLysLeu 1230
      |||||
2535 GAGTCCCTGACAA..... 2548
1231 AlaAsnLeuAsnArgLysTyrGluGluValCysGluGluValLeuHisAl 1247
      |||||
2549 ...ACTGTATCTCGAAATTTACATCATGTGATCATGCGATA..... 2587
1247 aLysLysLysGluLeuSerAlaLysAspGluLysGluLeuLeuHisPheS 1264
      |||||
2588 .....TCGAAAGATTAAGAGACATCTCTGTGACAT 2618
1264 ePILeGluGlnGluIleLysAspGlnGlnGluIleArgCysAspLysSerLeu 1280
      |||||
2619 CTGCCAAAAATCTTATCTACACATTCGCCAAAGCATATACAGTGAG 2668
1281 Thr..ThrIleThrGluLeuGlnArg..... 11 1290
      |||||
2669 ACACCAAGAAACCAAACTACAGCAAGAGAAACTGAAATATACCCAT 2718
      |||||
1290 eGlnGluSerAlaLysGln.....IleGluA 1299
      |||||
2719 TGAAGAAAGTAAAAAAGAGAAATGCGCTTGAAATTGATATTAAATT 2768
      |||||
1299 IaLysAspAsnLysIleThrGluLeuLeuAsnAspVal.....Glu 1312
      |||||
2769 CAGATAGCTGAGAACTACTGATCTTTGAGCATGTTTCAGAGAGAGAG 2818
      |||||
1313 ArgLeuLys 1315
      |||||
2819 ACATTGAA 2827
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2002, 16:31:41 ; Search time 39.27 Seconds

44160000 (sequences)  
2727.610 Million cell updates/sec

Title: US-09-750-590-2

Perfect score: 7039

Sequence: 1 MMSGSCAPRRKRNADNR.....DYQALLQIQHGLGVCSA 1401

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 7617452 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database: PIR-68

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No | Score | Query Match | Length | DB ID  | Description         |
|-----------|-------|-------------|--------|--------|---------------------|
| 1         | 640.5 | 9.1         | 1410   | A57013 | early endosome anti |
| 2         | 630   | 9.0         | 1939   | A46762 | myosin heavy chain  |
| 3         | 627   | 8.9         | 1940   | S04090 | myosin heavy chain  |
| 4         | 626   | 8.9         | 2663   | S28261 | centromere protein  |
| 5         | 625   | 8.9         | 2116   | S26655 | myosin heavy chain  |
| 6         | 621.5 | 8.8         | 1931   | A59234 | slow myosin heavy   |
| 7         | 618   | 8.9         | 1930   | S67553 | transport protein   |
| 8         | 618   | 8.7         | 2135   | A57013 | myosin beta heavy   |
| 9         | 610.5 | 8.7         | 2135   | A57013 | myosin heavy chain  |
| 10        | 610   | 8.7         | 1940   | A24932 | myosin heavy chain  |
| 11        | 608   | 8.6         | 1961   | A16131 | myosin heavy chain  |
| 12        | 606   | 8.6         | 1934   | A18133 | myosin heavy chain  |
| 13        | 606   | 8.6         | 1938   | A14944 | alpha cardiac myos  |
| 14        | 603   | 8.6         | 1935   | S06006 | myosin beta heavy   |
| 15        | 602   | 8.6         | 1935   | A59286 | myosin heavy chain  |
| 16        | 602   | 8.6         | 1935   | A49175 | myosin heavy chain  |
| 17        | 602   | 8.5         | 1932   | A18174 | myosin alpha heavy  |
| 18        | 600   | 8.5         | 1932   | A18174 | myosin heavy chain  |
| 19        | 596   | 8.5         | 1972   | A41604 | myosin heavy chain  |
| 20        | 595.5 | 8.5         | 1927   | A59236 | myosin heavy chain  |
| 21        | 594   | 8.4         | 3259   | A56539 | myosin heavy chain  |
| 22        | 592   | 8.4         | 1940   | A29330 | myosin heavy chain  |
| 23        | 591   | 8.4         | 1999   | S21801 | myosin heavy chain  |
| 24        | 589.5 | 8.4         | 3187   | JC5837 | myosin heavy chain  |
| 25        | 587.5 | 8.3         | 3225   | A15230 | myosin heavy chain  |
| 26        | 587.5 | 8.3         | 3225   | A15230 | myosin heavy chain  |
| 27        | 585.5 | 8.3         | 3253   | A15230 | myosin heavy chain  |
| 28        | 584.5 | 8.3         | 1427   | S22655 | myosin heavy chain  |
| 29        | 584.5 | 8.3         | 1979   | S03166 | myosin heavy chain  |

|    |       |     |      |        |                    |
|----|-------|-----|------|--------|--------------------|
| 30 | 582.5 | 8.3 | 1972 | JC5420 | smooth muscle myos |
| 31 | 582.5 | 8.2 | 1938 | JC5421 | smooth muscle myos |
| 32 | 577.5 | 8.2 | 1696 | JC5421 | microtubule bindin |
| 33 | 577.5 | 8.2 | 1957 | T18077 | probable myosin he |
| 34 | 576   | 8.2 | 1269 | P84730 | myosin heavy chain |
| 35 | 575   | 8.2 | 1369 | S13055 | myosin heavy chain |
| 36 | 574.5 | 8.2 | 1937 | S38173 | myosin heavy chain |
| 37 | 573   | 8.1 | 2007 | S38173 | myosin heavy chain |
| 38 | 572   | 8.1 | 1938 | A59283 | myosin heavy chain |
| 39 | 565.5 | 8.1 | 1837 | T41023 | probable nucleos p |
| 40 | 565.5 | 8.1 | 1837 | T41023 | myosin heavy chain |
| 41 | 568   | 8.1 | 2017 | A36016 | myosin heavy chain |
| 42 | 566   | 8.0 | 1976 | A59252 | myosin heavy chain |
| 43 | 565.5 | 8.0 | 1475 | T33318 | hypothetical prote |
| 44 | 564   | 8.0 | 2057 | S61477 | myosin II heavy ch |
| 45 | 563   | 8.0 | 1938 | JX0178 | myosin heavy chain |

## ALIGNMENTS

RESULT 1  
A57013  
early endosome antigen 1 - human  
N:Alternate names: endosome-associated protein  
C:Species: Homo sapiens (man)  
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999  
C:Accession: A57013, S44243, S44243, S44243, S44243, S44243, S44243, S44243, S44243, S44243  
J:Title: EBA1, an early endosome-associated protein. EBA1 is a conserved alpha-helical  
J:Hit: Chem 270.11503.1511.1995  
A:Reference number: A57013; MUID:95286647  
A:Accession: A57013  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1410 <RES>  
R:Scaling: 1.0  
R:Scoring: 1.0  
R:Submitted to the EMBL data library, April 1994  
A:Reference number: S44243  
A:Accession: S44243  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-254, 'C', 256-257, '10', 260-276, 'A', 278-283, 'A', 285-519, 'D', 521-574, 'BQ', 5  
A:Cross-references: EMBL:W78998; NID:9475933; PDB:1A55632.1; PDB:9475934  
C:Keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane prote

Query Match 9.1%; Score 640.5; DB 1; Length 1410;  
Best Local Similarity 21.7%; Pred. No. 5.9e-12;  
Matches 297; Conservative 263; Mismatches 440; Indels 369; Gaps 52;

OY 220 GDSHSY--YARIGNDITRLTKTASN--SMKGRH-----KKGSILQO-----261  
DB 67 GDSHSGSSNLAKRKDDVTLRLQVQDQASLKKRKYSEELKRELLKYGQDQAKK 126  
OY 262 -----RNLISQMLDEYNKRS--NORE-----HQLIDOLEIENEDKTRKRLKQOED 304  
DB 127 DGLVYDSSAEVQSLDEQLEAPQENEFIMQKMDLPEQMAQLATEINDIKSKYDKERSL 186  
OY 305 RILDKYKNGILOLNEEYVADLSE-----KEKLSILANKEKHESLRA 351  
DB 187 EAAEDKQYTRLEELNKEATYIDQLKELLQRLGRIGEVYVAKELVQYTLADKLTLENED 246  
OY 352 TLEALKSRKRYFSDHLDSG--SHPFKEDMLKKQGVWYTSQSTGTVHMSNRSL 408  
DB 247 ESKRKDESKRFESQYASSSARISQLSSE--LSKQPEYR-----YVVDLQKLT 293  
OY 409 RPL--ELALPNQSYSENEILKRELEMTFPCDSAKODRLQNELAKHVAEKKALALEC 466

```

Db 294 KSVNLTQKNO-TIFEN-LAKKE-----ODTFLK-----322
Oy 467 EKVVEDSDQIKOLDALDQOKRMVSESEKQVOMQWHTFLAKHNLTSDAATGNRL--523
Db 323 ---EKHNEISVSK---KNIOATVIMQDIDQDQOSRL-----SAGETSLIRHVE 366
Oy 524 -----MEBKDQIKDMKVYECASAEVKRLNIOKNE---MLVEERKDECKLMEB 572
Db 367 LSEKCAQVQKIKBELEVEETQVIMKALFEKQVQDQKREKQCHQIQDQSLNIOJSHKLLER 426
Oy 573 NKRIQLELMSCELEBERKQIKLTPEKQDQDQSLAKLSIPIKPEFNKSLSN--ELN- 629
Db 427 EROQEGALIGRLKEQDQSLSEKIMQKQOVDQDQLTL-----SRLEQDKKEVYNSTELQH 481
Oy 630 ---EKAK-----KLIDREVERSLNTRPLKRELEMLAKIAQD-----HV 667
Db 482 QLDITQKQNOHQBQALQOSTTAKRLKQNDQVQVIMQIGEKQDQIQDLEKQSLKSEKSNIS 541
Oy 668 KPEHPOQLKSHLEQKSGELKRIHTELQKQDQLEKQVGLMDKILQOQVNNLTPEMN 727
Db 542 LKREPRDLYAKIQAGEGETAV-LNQIQEKNTLQDEDTQV-----LENNKQKQSESHQ 593
Oy 728 VV-----LVYSEBKSKNDVTVDDLNKLSIDVTKKTEKYLEME---KILM- 770
Db 594 AQENMLHQVQKQKHLAADQVLSLETYSVLELNQSLKSEKQVQSDQIQDTAKTELLS 653
Oy 771 -----ENK-----SLKRVSLRETVPIPREKREKMAKISNTE 805
Db 654 ABAKTAQKQDQNLQDQVQNLQDQKQELKTTQDQDQVTKLQDKQKQSLQSHLKE 713
Oy 806 LKQDQLELNNKCGEQRKYSLSMEKNDKQKTSQVQVPTKHEIKTALSSTLKPNR 865
Db 714 YKEVYLSLEQKTELEQIKKLKLESDILEVKAKEQALQDQOQKLTQDLEKATTELSQ 773
Oy 866 L-----VQKKK---CEDIQEPTKQDNEIKLRLQENQVQKYSISREH 911
Db 774 LEMKEKYSSTRIDLOKKSPALSEIQVQKQREKQKQIKQKQEPQISQKTIQHEELNMR 833
Oy 912 EE-----KMSQLKSKMKVQDQSAELIAYKKGQGE---VTLHEE 949
Db 834 IQTTVTELOKVKKEKALMTLESTVQDKLSVDSILKNSKSEFEKQKQALIDLEKT 893
Oy 950 IAAQKREL-----OTIQEKTAKATPISLE---ECERKFAKTEKELQD---SQOQK 999
Db 894 CEKIQKQVQVQMENTLQKQELK---SLEKKAQSHQKLELMQDQIQDQNTQ 949
Oy 1000 NTSESE-----AKKQKQNDLKEKELTQDQD-----1027
Db 950 NKESEQDQLOGNINEKQSBQKQKQKQALQALQELKAVLQTELENNKIQDQVQAEILAA 1003
Oy 1028 ---KQKNVHNSYE-----TRALSQKTEEL--NR--1059
Db 1010 EKEKISVQNNYKESQDTPFKQDSDPYQGESEELATTRQDKLVYKELSLAQEDULISMRQ 1063
Oy 1054 ---QKQDLOKQTEAK-----KREKQVLEENKQKQSTIIAQTILQKQNV---PLEQV 1100
Db 1070 IQONQKQIDQELKATKATLEQDQSAKKQDQDQKQKQALQDQKESKLEKVLNKSQKLAET 1129
Oy 1101 ESUKSLSTLETLLEBLATKQRCYKQEQQVTVQJQMLENQKSSVLAELHQLKQKAPF 1160
Db 1130 EELKQKQEKETLQNLBELKSLKSLKSTETLQKQNDQQLQDQLEQKQASL---KAAV 1187
Oy 1161 KR---VQIIKASIRKREKESQNK---TEPYSKQJQSEIQNTQKQALKTLPEVQDVSQYKAT 1216
Db 1188 QKRNQOQIKQVQKKEEELKKEPILBEKAKHSEIKREKQVQKQKHEENR-----1236
Oy 1217 SQLEQTSQDNLKELKLANLRKYEEVCEVLAKKKELSAKQEKELHPSISEQITQDQ---1273
Db 1237 AKQIMQITLQMLNENLGVKQKQESQSRQVSLKQDQDQKQDQELQVLELVQNNQDQKRAL 1296
Oy 1274 EEC-----QKSLTITTELQRR-----IQESNQIQEADQKQITELN 1309

```

Db 1297 ERLKQSGEDELQTVKLEQRLQNTQVQGLQREMSQDQIHTQVQNL 1345

## RESULT 2

A46762 myosin alpha heavy chain, cardiac muscle - human

NContainers: myosin ATPase (EC 3.6.1.32)

C/Spectrum: Homo sapiens (man)

C/Date: 1993-09-10, Revision 31-DEC-1993, #test change 19-Jan-2001

C/Accession: A46762, A49354, S18830, B32562, B33835, B27856, A28908

R/Motusoka, R.; Beisels, K.W.; Furutani, M.; Arai, S.; Takeo, A.

Am. J. Med. Genet. 41: 537-547, 1991

A>Title: Complete sequence of human cardiac alpha-myosin heavy chain gene and amino

A/Reference number: A46762; MUID:92133665

A/Accession: A46762

A>Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Accession: A46762

A/Cross-references: DBJ:000943; NID:9219523; PID:BA00791.1; PID:9219524

A/Status: nucleic acid sequence not shown

A/Molecule type: RNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Accession: A49354

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

C:Genetics:  
 A:Gene: GDB:MYH6  
 A:Cross-references: GDB:120214; OMIM:160710  
 A:Map position: 14q11.2-14q13  
 C:Description: myosin heavy chain; myosin motor domain homology  
 F:187-168/Domain: myosin motor domain homology <score>  
 F:178-185/Region: nucleotide-binding site (NBS)  
 F:549-586/Region: actin binding site (ABS)  
 F:657-679/Region: actin binding site (ABS)  
 F:841-1939/Domain: coiled coil \*status predicted <COI>  
 F:841-1281/Region: 52  
 F:1282-1939/Region: light meromyosin  
 F:1282-1939/Region: light meromyosin  
 F:187-168/Domain: myosin motor domain homology  
 F:657-679/Region: actin binding site (ABS)  
 F:697-707/Active site: Cys \*status predicted

Query Match 9.0%; Score 630; DB 1; Length 1939;  
 Best Local Similarity 21.3%; Pred. No. 1; Gap 11;  
 Matches 291; Conservative 275; Mismatches 494; Indels 304; Gaps 55;

241 TASNRSNGRELKKGKGSIQ-----QRLNSQMLDEVNTR-----SNOHEQNIQD 285  
 Db 630 TDSGKSNKGGK-----KKGSPQVYSALHRELKMLKLNLTTHPHVVCITTEPFAQVMD 687  
 Qy 286 LTIENDLKK-----ERLR-----KIQOQORLL-----LDKWNQIQD 317  
 Db 688 NPLVMHQJLRCNGVLGRICRKGFPRIIVGDFRQRIILNPAIPEGQFIDSKGA--- 744  
 Qy 318 LMEVWVADLSESEKEL-----KSLIAAEKQIHESLFTT-----FALKSR 359  
 Db 745 -----KSLISLSDIDHNOYRFQHYVTFAGGLALEEMDRSLRITRIQAOARQOLMRE 802  
 Qy 360 FK-----YFESDHLGSSHFREKEDMLTKO-----GQMY 387  
 Db 803 FKKVIERBDALVIONTRAPGVYMMNPKLTYFKIRPLKSAETREKEMAKHEERQRIK 862  
 Qy 388 MTDSQSTCTQPMVHOSRNLK-----PLELALP-----NQASVSENEILKKELEAMPTTC 438  
 Db 863 ETLSEKARKELEKELKAVSLDEKNDQLQAOEDMLNDEERKCOLIKRNLQI----- 917  
 Qy 439 DSAR-----ODRLKQNEL-AKRYAVERKLLAEQVEESDQI KODLAKDQVQKRYE 493  
 Db 918 -EAKVEKNEIRLEDEEMNAELTKAKRLEDESLKXQID-----LETLAKYERKKA 972  
 Qy 494 SEQVAVQWQHTEFLAKELHTSDAQTGNRIIRMLEKDP-----LTKQWKK- 537  
 Db 973 TENKVNKLTREBAAGDITIAK--LTERKALBAHQADLDQAEEDKVTNYSKKKKE 1030  
 Qy 538 -----YEGASAEVGLANQIK-----QNF-MLVEERRODESKLM 570  
 Db 1031 QOVDDESLSEDEQKVKVMDLEAKRKLESGALQKESIMDLNDEKQLEKLEKLEKLEFIDN 1090  
 Qy 571 EENKRIQRELSCELEKREKRGKRLTPEMGOGLASALALSIPE-KFEKMSLILSEIN 629  
 Db 1091 QONKSKLEDE-QVALQALD--KIKMDQAVIIEELERLAKRYAKKELKLSIDLSSELE 1146  
 Qy 630 EKAKI-----LDEREYRSLNETRYLAKRELEKLANLQAQVFEEDQLKSRIDQ 681  
 Db 1147 ELSRLEPACGATSVQIBNNKKA-EQKMRDE--EAL-----QHRKALYALKSR 1196  
 Qy 682 KS-----GLKGRITELSKNOTLOKEIKYCDJMKLTQOVMNLTTEKKVNVLESEPK 1248  
 Db 1197 KHAQVADLEBDIDQARVOKRLEKSEKEL-----ELDDYSNMEDQ-IAKAKNLE 1307  
 Qy 738 KSIDVYVDQAL--KLSIDVHYKTEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 1307  
 Db 1249 KVSITLEDQANNEYVNLQEAQRLNDFTQBAKQLQDQBNELSLQKLEKLEKLEKLEKLEKLE 1307  
 Qy 795 EEMAKSKNTEELKQSLSEKAKGCEQKRYELKSENMNDLKKMKSHQVYVATEEELKTA 854  
 Db 1308 --LSTTQMDQELDKRQLEBGRK--AKNALAHMQASRHQC-DLIDQY--EETETAKAE 1358

Qy 855 LESTLTKTNEL-----VDKAKKEDINOEFVK-----IKDESELK-----RNLKN 896  
 Db 1359 LQVLSKANEVQACHTTKETIDALQPTTELEPAKKLQALQDQAEVAVNNAKCSLEK 1418  
 Qy 897 TQNVAVATYSIRREHEKNSGLKSMKKVQDDSAEILAKYKSOEIVTLTAETAAKRE 956  
 Db 1419 TKRHLDNEDIDLVDERNNAAAALDKQRFNFKILAAKMKQYEESSQSSLESSQNEAS 1478  
 Qy 957 LQTDICIKTAPLITSLSECKRRFATKELKQSLQSQOTQYNTSEBAKCKROENDL 1016  
 Db 1479 LST--ELFKANVSESLHLE--TKRENNKQLEISDQTLQDEGCGVNHLEKRNQL 1535  
 Qy 1017 KKEITLQDQDKNNHNSYETLEKLSKTE-----ELMQLKQDQYTKAKKE 1070  
 Db 1536 EVKLEQLQSLAEKVASLE--HEEKKILPQOLFQNOIKRTEKLEKLEDEKQAKRNHQ 1593  
 Qy 1071 KLVENAKQTSILAAQTLQKQHPVLEBOYESKLSSTETLTLEELTKQRYKREQQ 1130  
 Db 1594 RYVD-----SIQTSPLAEKNSVPLAEHLQVKEKPEKQNTIKSLREKVEKQNTKEPQVQSE 1643  
 Qy 1131 TYVQLRQMLEKQNSVPLAEHLQVKEKPEKQNTIKSLREKVEKQNTKEPQVQSE 1643  
 Db 1644 QVSIQSL--KQDQIQLDQVAVANDQLEKNIATV-----EBRNQLQDELEELAV 1693  
 Qy 1191 TQNTQALKLEETREYVDLSK-----YKATQSQDQIQLNLEKLANIKRYEVEE--- 1243  
 Db 1694 VEDTSS--KALADRELLETSSRQQLHSQNTSLNKKRKMDQLSLOLSEVAVQDECRN 1752  
 Qy 1244 -----VLHK-----KKELSAKDBELHLSIDEKIQOQERQNS-----LT 1281  
 Db 1753 AEKKAKKATITHHAAAEELKREQDQTSNHLERKKNMDQITKDQIRLDEARQIALGSKK 1812  
 Qy 1282 TITELRRIOESAKQILEAKDKITFLMDVRLKQALNGLSQLTGSGSPK--RQSQL 1338  
 Db 1813 QDQLEKLEAVLELSELEADQNRNAE--SKQMKSEKRIKELIYQTEDEKKNMLLQD 1869  
 Qy 1339 TDSLOQVRSLOQDQADQDEQVIAIR--THLSAAGQMD 1800  
 Db 1870 VDRQLQVAKYRQAELEQDQNTLKKPRYQNHLEDAEQRD 1913

RESULT 3  
 S04090 myosin heavy chain 3, skeletal muscle, embryonic - human  
 A:Cross-references: EMBL:U00001; GenBank:U00001; J:U00001  
 C:Database: S04090; Species: Homo sapiens (human)  
 C:Accession: S04090; S06146; S05444; S12460; S09333; S15082  
 R:ELIAR, M.; Steedman, H.H.; Sylvester, J.E.; Pertels, S.H.; Hubinstein, N.A.; Kelly, Nucleic Acids Res. 17, 3591-3592, 1989  
 A:Title: Nucleotide sequence of full length human embryonic myosin heavy chain cDNA.  
 A:Reference number: S04090; MID:89263803  
 A:Accession: S04090  
 A:Species: Homo sapiens (human)  
 A:Molecule type: mRNA  
 A:Residues: 1-1940 <EHL>  
 A:Cross-references: EMBL:X13988; NID:934843; PIDN:CAA3167.1; PID:934844  
 R:ELIAR, M.; Steedman, H.H.; Sylvester, J.E.; Pertels, S.H.; Wu, O.L.; Roychowdhury, P.B.S. Lett. 256, 21-28, 1989  
 A:Title: Human embryonic myosin heavy chain cDNA. Interspecies sequence conservation  
 A:Reference number: S06146; MID:90033298  
 A:Accession: S06146  
 A:Species: Homo sapiens (human)  
 A:Molecule type: mRNA  
 A:Residues: 774-1662, 'GT', 1665-1940 <EHL>  
 A:Cross-references: EMBL:X13100; NID:931143; PIDN:CAA31492.1; PID:931144  
 R:Kirsch-Mizrahi, I.; Travis, M.; Blau, H.; Levinand, L.A.  
 Nucleic Acids Res. 17, 6167-6179, 1989  
 A:Title: Expression and DNA sequence analysis of a human embryonic skeletal muscle m  
 A:Reference number: S05442; MID:8936648  
 A:Molecule type: DNA  
 A:Residues: 856-1390, 'KK', 1393-1940 <KAB>









Db 738 - ERLGSLDIIDHNYFGHVFYFAGLILGILEMBDRSLITRLOAQROQIMLRE 796  
 Oy 361 K-----YFSDHLAGSGSIFPKESDMLKGGCOMVYTD 390  
 Db 797 KLLERDALLVIGNMTRAMPQVKNPMKLYFRTKPLKSLAEFKEMOTKEEGHLE 856  
 Oy 391 SOTSOTGHVHNSR--SMRLP--ELALPNQASYE-----NELIKLEAENRFCD 439  
 Db 857 ALEKSEARKELEBERKYSMLQEKNDLDLOVQVQDQALADQEKNDQULINKTDO 910  
 Oy 440 SAR-----ODRLKQNEL-ARKVAECSKALCEVDESDROLOJEDKALDYKORWES 494  
 Db 911 EAKVEWTERLEEBEENMELTAKKRRLDEDSCELSKQIDID-----LELSLAKVEKHA 966  
 Oy 495 EGVYKQOQHFLALKEHLTSDAANGHRLBELKQID-----MVKRGS 540  
 Db 967 EKVYVNLTEBAGGDENIT--LTREKTLQDSHQDADLOADQEKVNTLAKVAKRQD 1024  
 Oy 541 ASAVGKLNQIKQNMELYEERK-----DSKLEENKIKLELSN 582  
 Db 1025 QADDESLTQOEKRTIMDLBRARKELEBQDLQAGSVADLENDQOELEKQIFELANT 1084  
 Oy 583 CELEBERKQKLTMEQOLQDLASR--LALSTPE-----KFNKKSLSNLEMKAKK 634  
 Db 1085 LAKTEBQALAAQOLKLEIQARLEEBLEBLETGKAEFTGKAPVRLSELQDLEETSR 1144  
 Oy 635 L-----IDYRETEPSLNETPRLKPELEENKIKLACIQVAFREHQLSKLEKSGEL 686  
 Db 1145 LEBAEGATSVOLTEKKRKA--EPQKRLRDEE--EATL-----QHEKTAALKR 1194  
 Oy 687 GRRTTELTKNQTLKELEKVDLNMKLQOVVNTKLEEMKNVPLKSEMKKSHDVYD 746  
 Db 1195 VAEISQDLQNMQVQKQLEKSESLTL--ELDVSSVNHQD--TKKANKLEKCKSTEDQ 1230  
 Oy 747 LKK-----KLSQVTKYKTEKLEMEKLKMSLSKVSNSLEYVTPERHEKEMALSN 803  
 Db 1231 MEHRKKLEESORTYDLOSTORAKLOTEHSLERQLEENF--ATL-----KQMLGRLTY 1303  
 Oy 804 TELKQOLSELNKKCEQDEK-----TYSLSMENDKTKMSHQVPAVKHEIKTALSTL 859  
 Db 1304 TO-----QLEDLKRLQLEEKAKKMLALHAGLSKQID--DLRLDY-----EEEMKAKLORAL 1356  
 Oy 860 DRYN-----RELVDYKRCEDINOETFKIKDENETLKRNLKENTONQY 901  
 Db 1357 SKANSSEVQWRTYETDQATQRTBELEKAEQALQDQEMEVAVNNAKSSDLETHRL 1416  
 Oy 902 KAEYISLREHEKMSGLKSMKYVQNSAELELAKKQSGEELVYALHEEIAQKREPTDQ 961  
 Db 1417 QNEEDLADYDERSMAAAALQKKQKQNDKTLSEBKQKKESSORTLELSQKARSLSST-- 1474  
 Oy 962 ECKTKAPATISLECEKRFKATEKLEKQESQOTQYVTESEBAKCKQKNDKJKEKL 1021  
 Db 1475 ELPLKMYVESLEHL--TFKREKNDQDELSLQDLOQAGSSHLELEVYKQDLEKL 1533  
 Oy 1022 TLQDQKRVNHTENSYETERALSRTKELENLQOLQYKTEKREKLEKLEKAKOTS 1081  
 Db 1534 ELQALAELEVASLEHE--EKKILRAQLEFN--QVY--AEVEKRLAKDEMEQSRKML 1586  
 Oy 1082 ETL--AAQTLQKQHVPLEGVSLSKLSGTLFTLEKLELTQKQCEKQDQATQOLQRMLE 1140  
 Db 1587 RYVQSLQSLDNESTRKNAQLKMSKQMLQENQDLSHANSRPAEAKQVAKALQVY-- 1645  
 Oy 1141 NOKKSSVLAELHLYKVAPEKQCIK-----ASLKEEESQONKSEKLOE 1190  
 Db 1646 --KQTOLODQAVPANEQDKENLAIYEBRRNLLQSELELTAWQDSRRAKTLKDEBLE 1703  
 Oy 1191 IQNTQKALKLETFREVVDSLKYVATQSQDLSQTOISLMEKLANMKRYE--VCEVYLHAR-- 1248  
 Db 1704 ASERVOLLHSQNTSLINQKKMEADISQLOTEVEBALQOCNNABEKAKKATIDAMMAE 1763  
 Oy 1249 --KEFLSADKEXLELSTISQELKQOEKSCS-----LTTTELQRIQESAQOIEA 1299  
 Db 1764 LKRBQDTSATLBNKKNQGVYDQALDDEABQALKQKQKQDQALQDEKVALEHELENEA 1823

Oy 1300 KDKRTLELNDYERLKAQNLNGLSQULYTGSGSPKRSQSLDLSQOQVRSIQOQADAPRO 1359  
 Db 1824 EQRNANESITQGLAKSERKVELSTYQTEDEKRNMYLQDLDVKLQKLVAKVRAQEADEQ 1883  
 Oy 1360 HQEYAIYFPHLLSMAQCHMDQENQEA 1366  
 Db 1884 ANSLNAKFR-----KQHELEDEAESRA 1905

RESULT 7  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A:Accession: F56753  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 1 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 567533  
 A:Accession: A38455  
 A:Species: Saccharomyces cerevisiae  
 A:Name: Nucleotide phosphatase 2 (Nucleotidase)  
 A:Date: 12-Jul-1996  
 A:Revision: 530782  
 A













A:Introns: 66/3; 114/3; 167/1; 176/2; 212/3; 265/1; 298/1; 332/3; 379/1; 418/3; 23/3; 1389/2; 1450/3; 1506/1; 1547/3; 1650/3; 1748/3; 1760/3; 1852/3; 1884/3; 1929/3  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 G:Keywords: actin binding; ATP; cardiac muscle; collod coil; heart; muscle; nucleotide b  
 F:87/65/Domain: myosin motor domain homology <MOT>  
 F:177-184/Region: nucleotide-binding motif A (P-loop)

Query Match 8.6% Score 606 DB 2: Length 1934;  
 Best Local Similarity 21.0% Pred. No. 8e-11;  
 Matches 285; Conservative 279; Mismatches 490; Indels 306; Gaps 53;

247 NGRRLKQKPSIQ-----ORNLQOMLDVNTK-----SNORHONQDLETENE 291  
 Db 631 DKQKAKKAGSEFQTVSYLREHKLNTLSTHPRHRCITLIPETKSGVONPVLVH 690  
 Oy 292 DLR-----EHLR-----KIQGQRL-----LDKYNGLDOLNEM 333  
 Db 691 QLRQNVGLSGIRICRGFPNRLLYGDFNQRYRILNPAALPGQFIDSKKA-----EKL 745  
 Oy 334 VADDESEKEL-----KSLAAEKQHEEL-RTTEALSKRRKFFPSHIGSSSIF 374  
 Db 746 SSLLDIDHNYKFGHTVFFKAGLLGLEKREHRELSRTITIDA-----QSGGLSMKF 800  
 Oy 375 -----KREDMLKQGMHTDSCTGTFVHMOSSMLRFL-----ELALPQA----- 819  
 Db 801 KILLERDLSLVIONMIRAPGVKWKPMKLVFKITFLKSLATEKEMAKMEFGRYVD 452  
 Oy 420 SYSENELTKLEEMRTF-----CDSANQRLKQNEL-- 480  
 Db 851 ALKSEARRELEEKVSLQEKNDLQVQAEODMLDAEERCQDLIKKIKOLEAKVE 920  
 Oy 453 -----AKHYACALALECERNEDSDQIQLEDAIDVQKRMTESEKVK 500  
 Db 921 MTRLEDEEEMAKELTKRKRLDEDSLLND-----LETLAKVEKMAKATKENVN 976  
 Oy 501 MQRHLLAKELHTSDAAGHRLMELEKQDLD-----MKRYEASAEVC 546  
 Db 977 LTRKVALDEIILAK-LTRKKALQEAHQOALDQDAEDVATVLTSSVKLE-----QYVD 1031  
 Oy 547 KLRNQINQNMILVEERDEGKLMENKRLQELSLCELEKRR----- 590  
 Db 1032 DLGSLDEQEKVNDLEAPKRLK-BQDLATQDE-STMDQOQDQDLERIKKDFELAL 1089  
 Oy 591 -----GRKLTMEQOLKSLAKSLASTPDE-KPMKSLISNLEKAKTL 635  
 Db 1090 NARIEDQALQSLOKKLKELOARIELEBELEMTAANAIVKLRDLSRLELEISRL 1149  
 Oy 636 -----IDVREYRSLSNETRPLKTELLENKAKLAGVAPREHODKRLDEKS--- 683  
 Db 1150 EAGAGATVQIEMKRRK-BQKRRKDE-EATL-----QHEATTAALRKIASDV 1199  
 Oy 684 GELGRTTELTKSMQLOKRELYCVLCKNLKLTQOVNKLTKEMVPLKSEPMKSHDVT 743  
 Db 1200 AELGQIDNLQVWQKLEKSEFRL-----ELDDVTSMDQI-LKANKLEKCRPL 1251  
 Oy 744 VDQNLK---KLSDTKRYTEKLEKEMKLKEMALSKNVASLETVLIPERHEKEMAK 800  
 Db 1232 EDDMHERSALETOBSVNDTQORAKLTQENGLSTSRQD-----EE--ALI 1297  
 Oy 801 SATTELK---KQSLSEKKGDEOK-----TYSLSMNDLAKTMESHQVPTVTEETIK 852  
 Db 1298 SOLTKGLVYTGQLEDLKQKLEBEVAKAKTALAHLSARHC-DLNEQI-----EETELK 1353  
 Oy 853 TALSSTLDQTNREL-----VDVKKCEDINQEVK-----IKQENELK-----RNL 894  
 Db 1354 AELQGVLSANSEVAQMTKYETDIALQRELEELAKKLLQNRJDAEAEVAEVMACSL 1413  
 Oy 895 EPTNOVAKETISREHEKSGLSKMKVQDASAILAKVKSQSEVTLIHEETIAQK 954  
 Db 1414 EKRIRQNEIEDLADVDSRANAAALDKQNFKEKILAMKQKTEESQSSQSEA 1473  
 Oy 955 RELDTQECIKIKIYATISLEPCBEKFKATEELKSQLSQSQQTQNTNSEPAKKQEMD 1014

Db 1474 RLISLT-ELFKLNAVEESLEHLE-TTRKRNKNLQGEISDLTFOGUSTGKSIHLEKRR 1530  
 Oy 1015 KLRKELITLQNDLQKDNVHEINSEYTRFSAKSTFEELNMQKLTQVYTAERKRLVE 1074  
 Db 1531 QLEAKREHLSALEEASLEHLE--GQNLKALQEFQDIAEIRKLEKDEMOAKR 1587  
 Oy 1075 ENAKQTSILAAQTLLQKQVHLEQVESLKSSGTTLETLEELTKQRCYERQOYVQ 1134  
 Db 1588 NHRVYDLSL-----QSLDIEFSSNALKRVRKQKQDQMDQDMLSHANRAAKQKQKS 1644  
 Oy 1135 LPMLEKQNKSSVPLAEHLYQKAEKEVGIITKSLRKEEESQKQTEYSQDSLEIOWT 1194  
 Db 1645 LQSLTL--KDTQILDQVANDDKENIALIV-----ERRNNILQAELEELHVAEQT 1694  
 Oy 1195 KQALKLETRREVYDLS-----KYKATKSDLETQISDLEKMLAMLRK 1236  
 Db 1695 ENE-KRLAQDELLETSERVOLLSQNTSLQNKRRMDLSQLOTEVEEAADVQCRMAEK 1753  
 Oy 1237 YEE-VCEVETLNRK--KKELSKQDEKELHPSIDQIKQOORCRMS-----LTTITE 1285  
 Db 1754 AKKATIDPANAAMELEKEDQPSAHLEMKRKNNQEQITKQILQIRLDAEQAIALKGGKQKQ 1813  
 Oy 1286 LQGRIOESAKQTEAKDNKITYELNDVYERLQKALNGSLQITGSGSPSEK--RQSQILSL 1442  
 Db 1814 LEARFLEMLELEQKRMNE--SVKQKSEERRKLELYQTEEDRKRLRLQDLYPKL 1870  
 Oy 1343 QOONRSLOQIADQDQNOQYVIAIIR--THLLSAQGHMD 1380  
 Db 1871 QLRVAKKQKQAEBAEQANRNLNLSKFRVYGHLEDAEERAD 1910

## RESULT 13

19464 alpha cardiac myosin heavy chain - mouse

C:Species: Mus musculus (house mouse)  
 C:Accession: U02301; Date of release: 02-Jul-1986; Year, change 02-Feb-2001  
 C:Accession: 19464; 19463; 19465;  
 R:Ouilin-Laguier, B. K.; Kennedy, J. E.; Mel, S. J.; Belajou, K. M.

A:Title: Characterization of the allelic differences in the mouse cardiac alpha-myosin  
 A:Reference number: A38207; MUID:92250040

A:Accession: 19464  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
 A:Cross-references: GB:U76601; NID:g191621; PIDN:AAA37162.1; PID:g191624  
 A:Accession: 19463  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
 A:Cross-references: GB:U76600; NID:g191621; PIDN:AAA37161.1; PID:g191622  
 A:Accession: 19462  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
 A:Cross-references: GB:U76599; NID:g191619; PIDN:AAA37160.1; PID:g191620  
 A:Accession: 19461  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
 A:Cross-references: GB:U76598; NID:g191617; PIDN:AAA37159.1; PID:g191618  
 A:Accession: 19460  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76597; NID:g191615; PIDN:AAA37158.1; PID:g191616  
 A:Accession: 19459  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76596; NID:g191613; PIDN:AAA37157.1; PID:g191614  
 A:Accession: 19458  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76595; NID:g191611; PIDN:AAA37155.1; PID:g191612  
 A:Accession: 19457  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76594; NID:g191609; PIDN:AAA37154.1; PID:g191610  
 A:Accession: 19456  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76593; NID:g191607; PIDN:AAA37152.1; PID:g191608  
 A:Accession: 19455  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76592; NID:g191605; PIDN:AAA37150.1; PID:g191606  
 A:Accession: 19454  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76591; NID:g191603; PIDN:AAA37148.1; PID:g191604  
 A:Accession: 19453  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Cross-references: GB:U76590; NID:g191601; PIDN:AAA37146.1; PID:g191602  
 A:Accession: 19452  
 A:Status: preliminary; translated from GB/EMBL/DBJ



F:548-585/Region: actin binding #status predicted  
 F:655-677/Region: actin binding #status predicted  
 F:839-1935/Domain: coiled coil #status predicted <C01>  
 F:839-1278/Region: 52  
 F:1280-1315/Region: light metomyosin  
 F:1280-1315/Region: light metomyosin (lys) #status predicted  
 F:184/Region: 184  
 F:695/705/Active site: Cys #status predicted

Query Match 8.6%: Score 603; DB 1; Length 1935;  
 Best Local Similarity 20.6%; Pred. No. 9, 8e-11;  
 Matches 282; Conservative 273; Mismatches 493; Indels 318; Gaps 51;

247 NGCELEKAKGSLQ-----QNLSDMLDEVNTR-----SNOREHNDIDLETEN 291  
 632 DKCGGAKKSSQVYSAHREKALREKALMLNLSSTPHFVACILINETSFGVNDPLVNH 691  
 292 DLK-----ELRLIOQ--EORILDVNGLOLOLNEWMYNDOLSEKELSLAKKX 344  
 DB 692 QLRNGVLEGRICRGFNRILVGFQNRILPALIPEGOFIRSGKGLLSGLDI 751  
 345 QHESLRTLEAKRFF-----YFSDHLSGSGFPRKEDMLKOCQWMTDSQCTSTGM 398  
 DB 752 DH-----NOYRFGTKYTFKGLDLLEKDEML-----SRIITRLQASRVS 795  
 399 PVHSGRSL-----NOYRFGTKYTFKGLDLLEKDEML-----SRIITRLQASRVS 795  
 DB 796 LSRHEKTLERDRLIIONNIRAPGVNMPMKLYFKILPLKSAFEKEMAKRKE 855  
 418 ----QASISENLAKELANRFF-----CDASAKODPLK 448  
 DB 856 RYKDALEKSAKRLKEKMYSLLOKNDLOLOVANQNLADERCDOLIKNKIOL 915  
 449 ONEL-----AKVACALALECEYKEDSDOLKOLEPDAKOVKRWES 494  
 DB 915 EAKYKEMTERLEDEEEMNNAELTKKRLDECSBELKGLDID--LEITLAVKERNAT 971  
 495 RYKQKQOTIFALKELTSDAOTGNHRLMEIKDKOL-----KRYVYG 540  
 DB 972 ENYKNTLFEEMAGDELIVK--LTKKALQENHQALDLOADEKVTITRAKVALLE 1028  
 541 ASAEVQIKRNOIKONELVEPKRQDGLMEKNRLOKELSGEELPERK----- 590  
 DB 1029 --QOVDLEGLSDDKKVRMDLBRARFL--EDLKLIT--SINDLENNKOQDERLKKD 1084  
 591 -----GRKLTEMGQLOKLSAKLALSLIPE--KEKNNKSLISLSEIN 629  
 DB 1085 FELMALNKRLEDOAGLOKRLKALQANLITELFEELKARPAKVEKELSLDSLELE 1144  
 630 EKAKRL-----IDREYERSLNPRLRRLREKLKALALQVKNRFEHOLKSLRDO 661  
 DB 1145 EISERLEAGATSVQIEKMKREA--EFQKMRDLE--EATL-----QHEATMAALX 1194  
 682 KS-----GELCKRITELTSKNOTLOKELEKXCDKLLTOOVNLTTEKKNVPLVSEPK 737  
 DB 1195 KHNADVYADLQK--KLSDVTHKYTEKKLEKMLKNSANLSSTLETVLEPPEHKE 794  
 738 KSNADVYADLQK--KLSDVTHKYTEKKLEKMLKNSANLSSTLETVLEPPEHKE 794  
 DB 1247 KQORLEDOENHRSKAEFTORSVNDLTROKARLOTENGELSRLD-----EK 1294  
 795 EMMALKSNITELK--KQISELNNKGGEOQK--IYSJSENNDKTKTSNQVVPK 846  
 DB 1295 E--ALISGLRSGKLTTLQOLDLKQLEBEVANAALMAHLSAHHC--DLIRDOY--- 1348  
 847 THEEKRTALSTLQTNEL-----DYKKKCGDINQGV-----IKDENILK--- 931  
 DB 1349 EETFAKKELOVRLSKANSEVQWPKRYKEDALOTETEELEKAKKLAKROLODEANVAVN 1408  
 892 ---RNLGNTOVYKAEYISLREHEKKSGLFSGKSNKVVQDANSJALIAKYSQSEIYTLHE 948  
 DB 1409 AKCSLSLTKRHLQNEIDLVADVERSSANAALDKKORNDKILIVEMOKYEESOSELE 1468

949 EIAKORRELTIOETKILKAPITISLECEKRFVATEKLEKOLSOOTQVNTSEBANK 1008  
 DB 1469 SSKORNSLST--ELRLKNNRESLEHLE--TFKRNKRLQDEISLITEDOLSTGKSH 1525  
 1009 KQOENKTKREKLTLLQOLKQKVMYEMSVTEKALSRKTEMLQOLDKOTYKAKE 1068  
 DB 1526 LKTRKOLEKTELOSALEAASLEH--EKLIRKOLENOIKVIEKRLAKBDE 1582  
 1065 KEKLEENAKQTSLELAOTLLQKQVLEJOVESIKKLSSTETLKEKRLKORCYKE 1128  
 DB 1583 MEQKRNKLNVDSIL--OTSLEKRSNKLALNKVKKKEDDLMELIOTISARNAVA 1639  
 1129 QOVVOTLOKLEKNNKSSVPLAFLQVLEKFEKVICIKASLEKEESQNTKERSKLO 1188  
 DB 1640 OKOVKSLOSL--KQTOLODOAVRANOLKENTILV-----ERNILLOLELEK 1689  
 1189 SETIOTKQALKLETRPVDS-----RYKATSDLETOISDLNKL 1230  
 DB 1690 ANVANKYEE--VCEPVILAK--KXELSAKDEKELSLSEIDRIKOOEKDSS-- 1279  
 1231 ANVANKYEE--VCEPVILAK--KXELSAKDEKELSLSEIDRIKOOEKDSS-- 1279  
 DB 1749 RNEEKAKKATIDAAVMALEKLEDOVSALHREKNNKEOTINDORLDEKADALAG 1808  
 1280 LTTTELORIOSAKOTJEAKNDKITELNDVLEKQALNLSLTYGSGSPSK--RQS 1336  
 DB 1809 KQOLKLEARVLELELEKQARNAE--SYKQKRSERIRKELTYOTEDEKRNKLRLQ 1865  
 1337 QLDSLOOQVRSLOOLDADPROQEVILAYE--TILSLAOCIND 1380  
 DB 1866 DLVDKLOLKVKAYRQAEKAEQDQNTNLSFKRYQVHELDENBERAD 1911

RESULT 15  
 15026 heavy chain beta chain, cardiac - pig  
 A:Accession: A59286  
 C:Species: Sus scrofa domestica (pig)  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 08-Sep-2000  
 C:Accession: A59286  
 R:Ko, Y.L.  
 submitted to Genbank, October 1996  
 A:Accession: A59286  
 A:Accession: A59286  
 A:Notes: preliminary; not compared with conceptual translation  
 A:Residues: 1-1935 <R01>  
 A:Cross-references: GR:075316; NID:q1698994; PION:AB3730.1; P1D:q168995  
 A:Experimental source: strain domestic  
 C:Keywords: cardiac muscle; heart  
 F:86-766/Domain: myosin motor domain homology <M0>

Query Match 8.6%: Score 602; DB 2; Length 1935;  
 Best Local Similarity 22.0%; Pred. No. 1e-10;  
 Matches 266; Conservative 260; Mismatches 451; Indels 233; Gaps 49;

261 QRLSDMLDEVNTRKSNOREHNDIDLETENSDKRLKLOOQRIILDKRNGLOLQNL 320  
 DB 846 EKAKTAKREDEG--RLKEALEKSSARRELEKRWVSLDEK--NLDLOVQK 893  
 321 EYKVAWDALESSEKRLKSL-----LAAKQHEESLRTIALKSRRFYESDHLSGSHF 374  
 DB 894 E--QDNLDSERCDOLIKNKIOLFAKYKEMERLEDEEEMNA----- 934  
 375 REEDMLKQGOVYDSQCTSGCHVQMSKMLRLELALPNOASVSENLILKLEELAM 434  
 DB 935 ---ELTKKRVN---EDCS-----ELKROIDDELTLAK-----VEKERNAT 971  
 435 RFDQSKODPLALQNLQELAKVAESKALALCEKRVKEDSDQIKOLEDAKDVORWYES 494  
 DB 972 EKRYKNTLEPMAGDELIALTKKKAAL-----QANQO-----ALDLO--- 1013

```

QY 495 BGVYKONQOTIFLAKTELTSOATNGHMEELKODKMKVYBS-----ASAEVCKLBN 550
Db 1014 BGVYKNTLTKAKVLEQHV--DOLAGSLRQKVMRLERAKRKLEBDLKLTOESIMDLN 1071
QY 551 QIKONEMLVEEFRNDECKLMEBNRRLQKELSM--CELEKREKNGRKLTMEGOLKIDSAKLA 609
Db 1072 DKQO---LDERLKKDFELNALNARLEBDOLQSLO---KKLEQLARLEBELLE 1123
QY 610 LSIPLAC--KREKMSLSNLSNEKAKTL-----IDYREKFRSJMTRPKRELENK 660
Db 1124 AERTAAKAVKELSLDSLELEISERLEEGAGTSVQIDMKKREA--EFOKMRDLE--E 1180
QY 661 ARLAOHVFEHEBOLKSRLEOKS---CELOKRIETLSKNOTLOKEIEKYCLDKHKLTO 716
Db 1181 ARL-----OHEATAALAKRKHADSVALEGEOIDLONRYOKLEKESKFXL----- 1226
QY 717 OYNNLTTEKNVPLKASEKKSHDYIYVDLKN---KLSDVTKTEKKLEMEKLYMNA 773
Db 1227 ELDVOTSNKEDJ--IAKANLBNKCTLEDONMHEBSAKETROPSVNDLTQBAKLOTENG 1285
QY 774 SLSKNVSRLETVFIPPERHEKKPMALKSNITELK---KQLSBELNKKGEDEK---LY 825
Db 1286 ELSRQD-----EKE--ALISOLTRCKLTYYTOLEDLKQLEEVANAKNALAH 1331
QY 826 SLMSENN--DIKRTSHQVYPVKTHERIKTALSSTLQKTNREL-----VDKAKCED 875
Db 1332 ALQSAARHADDL--LRQY---BEETETALQRYVLSANSVYAKORTKYTEIDALORTBE 1385
QY 876 INOEFYK-----IKDENELK-----RLNLTQNOVMAEYISLREHEKMSGKRSMKVY 925
Db 1386 LFEAKKKKLQLODAEFAVAVNAKCSLEKTKHRLQNETEDLAVDVERNAALAAIDKK 1445
QY 926 QDNBAELIAXKKSQOEIVTLHEBIAQAKRELDTIOECIKLYAPILISLECEKFKATE 985
Db 1446 OKNRDOKILAEKORTESOSLESSEOKENSLST--ELFKLNAYESDENLITS--KREN 1502
QY 986 KETLEKLSOOTOKYNTSPBEAKKQKQONDKLKEKITLOKLDKXNHTENSYETERALS 1045
Db 1503 KMLQETISDLEOLSSQGTIHLELEYRQLEAEKLELQSALESSEASLEHE---EGRIL 1559
QY 1046 RKTTELNRQLDLQKYTEAKKEKXLYEENAKOTSSELLAOTLLQKQVPLEDYESIKK 1105
Db 1560 RQOLEFNOIKAEKMERLEKDEBMOAKRNHLRYDSL---QTSIDATERNRNALRYAK 1616
QY 1106 SLSTGTETLXELKTKORCYEKEDQYVTOLOHLEKONKSNVPLAEHLQVKAPEKVEGI 1165
Db 1617 KMEQDLNMEIOLSHANRMAEAOVKYKLSQSL---KDTQIOLDVAVRANDLKENIAI 1673
QY 1166 IKASIRKKEESONKTEVSUKLOSRIQNTQKALKLETRREVYDS----- 1210
Db 1674 V-----ERRNNLLOALEBELRAVAEOTERS--RKLDAQELLETSERVOLLSQNTSLIN 1725
QY 1211 ---KRYATSDLEFJOISLMEKLIANRKYER--VCEEYLAHK--KRELSAKDEKELLIFS 1264
Db 1726 OKKKMEADLSOLOFEVAEVOESRMAEPEKAKKITDAAMAEBELKKEODTSAHLEPMKKN 1785
QY 1265 IDEKIXDOQERCDKS-----LTTTELQRIQESAKOIEAKDNKITELLDYERLKO 1316
Db 1786 MROTIKDLHRIDRAEOLAKGKKQOLAEVARELELAEADKRNAE---SVAGKRR 1842
QY 1317 ALNGLSTLYTSSGSPK---ROSOLISLOOQVRSILOOLADAPROHOEYAIAYR--THL 1371
Db 1843 SBRRIKELTYOTEDRKMLRLQDLYDKLQKVAAYNRQAEBEQAQNTNISKFRVOHE 1902
QY 1372 LSAAGCHMD 1380
Db 1903 LDFAEERAD 1911

```

Thu Apr 11 07:43:35 2002

us-09-750-590-2.std.rpr

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2002, 18:08:47 : Search time 28.76 Seconds

(Without alignment cell updates/sec  
1766.072 Million cell)

Title: US-09-750-590-2

RefSeq score: 1 M5CSCFSCAFKRNQADMK.....DVALLIQIIMRGGLVCSA 1401

Sequence:

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 3666427 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SWISSPROT\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Match | Length | DB ID         | Description         |
|------------|-------|-------|--------|---------------|---------------------|
| 1          | 627   | 8.9   | 1930   | 1 MYH4_HUMAN  | P13533 homo sapien  |
| 2          | 627   | 8.9   | 2230   | 1 MYH3_HUMAN  | Q13439 homo sapien  |
| 3          | 626   | 8.9   | 2663   | 1 CENH3_HUMAN | P02754 homo sapien  |
| 4          | 625   | 8.9   | 2116   | 1 MYH2_HUMAN  | P13538 homo sapien  |
| 5          | 623   | 8.9   | 1940   | 1 MYH3_HUMAN  | P13538 homo sapien  |
| 6          | 618   | 8.8   | 1790   | 1 USO1_YEAST  | P25366 saccharomyc  |
| 7          | 610.5 | 8.7   | 1935   | 1 MYH7_HUMAN  | P12883 homo sapien  |
| 8          | 610   | 8.7   | 1940   | 1 MYH3_HUMAN  | P13538 homo sapien  |
| 9          | 610   | 8.7   | 1940   | 1 MYH3_HUMAN  | P13538 homo sapien  |
| 10         | 606   | 8.6   | 1934   | 1 MYH6_HUMAN  | Q02556 mus musculu  |
| 11         | 606   | 8.6   | 1961   | 1 MYH9_HUMAN  | Q02556 mus musculu  |
| 12         | 604.5 | 8.6   | 1960   | 1 MYH9_HUMAN  | Q02556 mus musculu  |
| 13         | 603.5 | 8.6   | 1938   | 1 MYH9_HUMAN  | Q02556 mus musculu  |
| 14         | 603   | 8.6   | 1935   | 1 MYH7_HUMAN  | P12883 homo sapien  |
| 15         | 602   | 8.6   | 1935   | 1 MYH7_HUMAN  | P12883 homo sapien  |
| 16         | 602   | 8.6   | 1935   | 1 MYH7_HUMAN  | P12883 homo sapien  |
| 17         | 600.5 | 8.5   | 2022   | 1 MYH6_HUMAN  | P13538 homo sapien  |
| 18         | 600.5 | 8.5   | 2022   | 1 MYH6_HUMAN  | P13538 homo sapien  |
| 19         | 600   | 8.5   | 1938   | 1 MYH9_HUMAN  | Q02556 mus musculu  |
| 20         | 596.5 | 8.5   | 1938   | 1 MYH9_HUMAN  | Q02556 mus musculu  |
| 21         | 596   | 8.5   | 1938   | 1 MYH9_HUMAN  | Q02556 mus musculu  |
| 22         | 591   | 8.4   | 1940   | 1 MYH3_HUMAN  | P13538 homo sapien  |
| 23         | 590   | 8.4   | 1940   | 1 MYH3_HUMAN  | P13538 homo sapien  |
| 24         | 584.5 | 8.3   | 1427   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 25         | 582.5 | 8.3   | 1427   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 26         | 582.5 | 8.3   | 1427   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 27         | 580   | 8.2   | 1084   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 28         | 578   | 8.2   | 1941   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 29         | 577.5 | 8.2   | 1957   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 30         | 576   | 8.2   | 1937   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 31         | 574.5 | 8.2   | 1875   | 1 MYH8_HUMAN  | P10587 gallus galli |
| 32         | 573   | 8.1   | 102    | 1 MYH8_HUMAN  | P10587 gallus galli |
| 33         | 573   | 8.1   | 1939   | 1 MYH8_HUMAN  | P10587 gallus galli |

| RESULT     | 1                                                                        | STANDARD  | PRT                    | 1939 AA. |
|------------|--------------------------------------------------------------------------|-----------|------------------------|----------|
| MYH4_HUMAN |                                                                          |           |                        |          |
| AC         | P13533                                                                   | Q13439    | Q14906                 | Q14907   |
| DT         | 01-JAN-1990                                                              | (Rel. 13) | Created                |          |
| DT         | 15-JUL-1998                                                              | (Rel. 36) | Last sequence update   |          |
| DT         | 20-AUG-2001                                                              | (Rel. 40) | Last annotation update |          |
| DE         | MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM (MYHC-ALPHA).           |           |                        |          |
| GN         | MYH4 OR MYH4C.                                                           |           |                        |          |
| OS         | Homo sapiens (Human).                                                    |           |                        |          |
| OC         | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;                  |           |                        |          |
| OC         | Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.                 |           |                        |          |
| NC         | NCBI_Taxid=9606;                                                         |           |                        |          |
| LN         | [1]                                                                      |           |                        |          |
| FT         | SEQUENCE FROM N.A.                                                       |           |                        |          |
| FT         | MEDLINE=92133665; PubMed=1776652;                                        |           |                        |          |
| FT         | Matsuno K., Beisel K.W., Furutani M., Arai S., Takao A.;                 |           |                        |          |
| FT         | Complete sequence of human cardiac alpha-myosin heavy chain gene and     |           |                        |          |
| FT         | amino acid comparison to other myosins based on structural and           |           |                        |          |
| FT         | functional differences. J. Biol. Chem. 268:11911-11919.                  |           |                        |          |
| FT         | A. V. Med. Genet. 41:537-547(1991).                                      |           |                        |          |
| FT         | [2]                                                                      |           |                        |          |
| FT         | SEQUENCE FROM N.A.                                                       |           |                        |          |
| FT         | MEDLINE=94140346; PubMed=8307559;                                        |           |                        |          |
| FT         | Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;                  |           |                        |          |
| FT         | Structural organization of the human cardiac alpha-myosin heavy          |           |                        |          |
| FT         | chain gene (MYH4). J. Biol. Chem. 268:11911-11919.                       |           |                        |          |
| FT         | Genomics 18:505-509(1993).                                               |           |                        |          |
| FT         | [3]                                                                      |           |                        |          |
| FT         | SEQUENCE OF 1-177 AND 1551-1939 FROM N.A.                                |           |                        |          |
| FT         | MEDLINE=98264452; PubMed=2726733;                                        |           |                        |          |
| FT         | Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;            |           |                        |          |
| FT         | Characterization of human cardiac myosin heavy chain genes. J.           |           |                        |          |
| FT         | Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).                        |           |                        |          |
| FT         | [4]                                                                      |           |                        |          |
| FT         | ERRATUM: Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;   |           |                        |          |
| FT         | Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;            |           |                        |          |
| FT         | Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).                        |           |                        |          |
| FT         | [5]                                                                      |           |                        |          |
| FT         | SEQUENCE OF 1407-1939 FROM N.A.                                          |           |                        |          |
| FT         | MEDLINE=98299163; PubMed=2969919;                                        |           |                        |          |
| FT         | Kutabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;         |           |                        |          |
| FT         | Molecular cloning and characterization of human cardiac alpha-           |           |                        |          |
| FT         | and beta-form myosin heavy chain complementary DNA clones. Regulation of |           |                        |          |
| FT         | expression during development and pressure overload in human             |           |                        |          |
| FT         | cardiac myocytes. J. Biol. Chem. 268:524-531(1993).                      |           |                        |          |
| FT         | [6]                                                                      |           |                        |          |
| FT         | FUNCTION: MUSCLE CONTRACTION.                                            |           |                        |          |
| FT         | HEAVY CHAIN SUBUNIT (MYC). 2 ALKALI LIGHT CHAIN SUBUNIT (MYL).           |           |                        |          |
| FT         | AND 2 REGULATOR LIGHT CHAIN SUBUNIT (MYL-2).                             |           |                        |          |
| FT         | LOCUS: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEITIVE, SHOWING            |           |                        |          |
| FT         | CHARACTERISTIC REPEAT PATTERNS COMPOSED OF 4 HEPTAPEPTIDES,              |           |                        |          |
| FT         | CHARACTERISTIC REPEAT PATTERNS COMPOSED OF 4 HEPTAPEPTIDES,              |           |                        |          |
| FT         | -1- PTH: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY          |           |                        |          |

## ALIGNMENTS









OY 281 ON-10DLEFENEDLKERLKOORLILDOXVNOLOLNEPVM-ADOLSE----- 331  
 DB 899 ROSLOJTEREKLTITKELQOTLEEVKTLIOEDOLK-OLQESLOJERDOLKSDIHOTVA 957  
 OY 332 -----KERKSLAKKEHESLRTITALKSFYFESHDGSGSHF----- 374  
 DB 958 NMIDJODOLNLESJ-KOHOE--TITLAKSI-----SEKVSNNLMEENTGETKIDERO 1009  
 OY 375 -----KREDMLKOGAMWTDGCTGCPHWSRMLRL-----ELALPNASYS 423  
 DB 1010 OKMVGIDKODLBAKNTQTLTADVYKNE-----TIDQOKRIFSLQEKNELOQOMLESYAN 1065  
 OY 424 NEILKER-----EAMRTFDSAKODBLJLONELAKVACAAALACERKYE- 471  
 DB 1066 KQOLTDJLAKENITPITNOBELKIDDELKQOOLVAGKEMAJAKKEESLRTICORLAEV 1125  
 OY 472 -----OSEBOIKOEDALDVOXKMESEKYOQHOTFLAKLHLJSDAONTRHAME 526  
 DB 1126 BEKKEKESQOLOKXOOLINTVOEPENSEOKKINE-----IN 1162  
 OY 527 LKPOLDKKVVYEGASAEVQKLPBQKONELVEFFRDEKLMENKRIQKESLMECE 586  
 DB 1163 LKNEILKELTILHETTERKELJAKLNN--YEEVK-----STIKREYVLEKLOKSEFE 1215  
 OY 587 BEKKSOKLTMEG-----OLDKSLALSTRAKFEFNKSKLSLNLMEKAKKIDOVERE 643  
 DB 1216 ROHLRGYIREITATQOTKE-ELKIA-HILKEHJETIDELARSVEKTAQILINQO----- 1269  
 OY 644 KSLNETRKLRELENKLAQHYKEEHRQK-----SRLEQKSGELKRTITELTS 695  
 DB 1270 -----DLKSHFTLODET-FLVHEEDELPLVKKVSEFOYTNELBELLEOSTT 1317  
 OY 696 KNOT-----LOKEITKYOVDNELLTOO--VNNITKKNPPLKASBEKKSHDVYDOLAKKL 751  
 DB 1318 KDSITLAIEMERLAKNEKFOESQEZIKSLTERDNK-TIKAKLVHNOQKELHETL 1376  
 OY 752 SDVTHKVT--EKLEMEKLMENASLKNVSRLETVIPPE-----PHEKSMALKSVITE 805  
 DB 1377 AKTOESQKODOSLMMK-----KODETTVIVSEMO-FKFKPSALALITETIMELSGRIQ 1432  
 OY 806 LKQSLSELNKKGEOKATYLSLSPNDNKTMSHOV-VKHEITRA----- 854  
 DB 1433 SHDEKSYAKENDOLQLOEVLOSQDOLKENTIKYANHLLEPBLVAKHCKLEBET 1492  
 OY 855 -----LSSTLDNRELVYVKKCDIN-----OPKVIKIDENETIK-RULN 896  
 DB 1499 INELRVASISKEFETSTIKOLAPINDKONKIOEITVEKOLINTKOISVEOVANNELQ 1552  
 OY 897 TONUYAVYSISREHEKMSCL-----KSKKKVO-----DN 928  
 DB 1553 FKBEHRKAPDALSIESKLETLNRLQOSQEOIOMIKKEKESKALALQERDOLAKN 1612  
 OY 929 SAEILAKKRSOE-----EYVHEBELAKOKKRELDIO-CEIKKY 968  
 DB 1613 TRELVAKKRSOEKOFIKNATVNETOKEKEIEMLKPOEPQKMLNENIETNRLQ 1672  
 OY 969 APIISLE-----CEKFAKTELAEOLSOO----- 995  
 DB 1673 ILHENLEPMHMYTERDRLBSYEFLLAKERDOLKEMENRETTIDOLEKQELIYVHHLKE 1732  
 OY 996 -----TOKVYSEKAKKQKENDKLKELITLONKODKKNVIENSUYET----- 1040  
 DB 1733 HOETIDKRLQVISEKTEKNSMOKDLENSMDLAADOLKIDEBELRIAMHLETCOEFIDK 1792  
 OY 1041 -EVALSRETEENLOKO-----LQATTEAKREKEL-----VEENAKOTSLEIA 1086  
 DB 1793 LRLVISEKTKNSMOKDLENSMAKQLOKLAENQOLITLAKDVOKOYKVSPOML 1852  
 OY 1087 QTLLOKHVLEBOVE-----SLAKSISGTEITLAKELKVT--ORORYEKEDVOYSLPOML 1139  
 DB 1853 KKOIKOOSLITSLKLETLNLAOLBHEKLEBKSVKMERDNLARVEFTLKERQOKKESL 1912

OY 1140 ENOKNSVPLAEHLOVEAFKFEVIGKISLAKKEEBSOONTEESKLOEYONVOKALK 1199  
 DB 1913 CERNARDELTOELKTRMLKREKREYVNDLNRIS--TTTOISIDOLDKSDRELO 1969  
 OY 1200 KLETRVYVDSKY-----AKRSOLETOISLNE-----KLAN-LMKRY 1237  
 DB 1970 K-----KIOELOKKELLOLVAREOVYNNKHTIKENDOLKKEFERNVCKCCHONQUTKL 2025  
 OY 1238 EENCEV-LANK-----KESLARD-----KELHLRST 1266  
 DB 2026 HESKELEIVYAKENDELRIKESLKKERDOFTATLRPMADRONHOKVKEKRLISDOO 2085  
 OY 1267 OXKIOEORCDKSLTTITLORLORIKESLOIDAKONKITELDVNE-----RKQALNG 1320  
 DB 2086 HLESIRKESR-----IKELKATSPMDHYCENLBSLDEKLEIFHRMKIKLVYS 2141  
 OY 1321 LSOITYGSGSPKRSQSLDSLOOVNRSLOOOLMDNRQHOEVYATRTHLMSAACHD 1380  
 DB 2142 VTKI-----KEBOHCINPFDMF-----IDEVOKELLIRI-OHL-----OOD 2180  
 OY 1381 EDVQALLOIOMRQL 1397  
 DB 2181 CDVPSHELRLKLNQNN 2197

## RESULT 4

MS2\_DICT STANDARD: PRT: 2116 AA.

AD MS2\_DICT  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MYOSIN II HEAVY CHAIN, NON MUSCLE.  
 GN MHCA.  
 OS Dictyostelium discoideum (slime mold).  
 OC Bacteria; Mycetozoa; Dictyostelida; Dictyostelium.  
 NC NC\_017410-44689;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE-87092266; PubMed-3540939;  
 RA Warlick H.M., de Lozanne A., Leinwand L.A., Spudis J.A.;  
 RA Conserved protein domains in a myosin heavy chain gene from  
 RA Dictyostelium discoideum.\*  
 RT Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).  
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.  
 RC STRAIN-AX2.  
 RA MEDLINE-90353583; PubMed-2387409;  
 RA Lueck-Vielmeier D., Schleicher M., Grabert B., Wippler J.,  
 RA Gerlsch G.;  
 RA Replacement of threonine residues by serine and alanine in a  
 RA phosphorylatable heavy chain fragment of Dictyostelium myosin II.\*  
 RT FEBS Lett. 269:239-243(1990).  
 RP PHOSPHORYLATION SITES.  
 RA MEDLINE-88112226; PubMed-2828113;  
 RA Maqle G., Noegel A., Scheel J., Gerlsch G.;  
 RA Phosphorylation of threonine residues on cloned fragments of the  
 RA Dictyostelium myosin heavy chain.\*  
 RT FEBS Lett. 227:71-75(1988).  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.  
 RA MEDLINE-95343063; PubMed-8619795;  
 RA Fisher A.J., Smith C., Thoden J.B., Smith R., Sutcliffe H.M.,  
 RA Rayment I.J.;  
 RA X-ray structures of the myosin motor domain of Dictyostelium  
 RA discoideum complexed with MgADP, BeFx and MgADP-ALF4.\*  
 RT Biochemistry 34:8960-8972(1995).  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.  
 RA MEDLINE-95343067; PubMed-7619796;  
 RA Smith C.;  
 RA X-ray structure of the magnesium(II) pyrophosphate complex of the



```

1469 EADRCNSAIFAKKATSALESLEKDEIDAMNANKAKERKSELEVALEVALELEKSG 1528
1470 OYKAEVISELHEHEKMSGLR-----SKKRVQDNASGELLAKYKSCSEIYV 945
1471 YVWEVI--KKNMEIDDLKARLDRETSRKSDCKKNTKOPALALAEVAGREVV 1586
1472 -----LHEIIMAKRELEDTQECIKIKAPILSEEEERK-----TA 983
1473 IDRLKKEEDIDTIDSTOLDT-----EKRSLIKERKSKKKLEOTLAERMAEESMA 1639
1474 TEKLEKLEQSLQOOR-----YNTSEERKAKOENDIKKKEILLLOKDKIKRV 1032
1475 ADEELIKKMOVOWEDELADOLSEBALALAKSEKIKSLAEVDEVKO-----LEBEL 1692
1476 HENKSEFETASLKRTEELNOLKJLDLXATKAKKEKEL-----VEKNKQTSFILAOL 1089
1477 AKDVLAKKRALEVALEVEVROQLFEEDSESELEKSRRLTTEYEDIKKVALEQENK 1752
1478 LONOHVP--BOVESLKSLSGTETLEKLEKLEKORCEEQOYVTOLOKLEKNSASP 1148
1479 LDKKKKLLDVLDTLKKKLEDEKKKIKLMSRRKSLSENDPFLAKLAEVKNRSRE-- 1810
1480 LAHLOVLEAFKREYVIGIYASLEKFEESQNKTEPYSKLOEYONTOALKLETRVND 1208
1481 -----KDRKRYEKDKLQTKYKND-----EAKKQTELOAKLEQIDELSKKE 1856
1482 LSKKATKSP-----LETOISD-----ANKLANKRYEEVCEVLEAKK 1249
1483 OENAKKTQAKSKRTLEGEIDNLKADIDEGKIKKLEKLEKLELELELELELELEAD 1916
1484 KESLAKDEKLEHSTIGETIKOQENCK-KSLT-----YTLEQRIQICE----- 1292
1485 SKSAPQSKRLVELELEDRNKLQKIDAKLEIADKAKSLQRLVAKGLEESKINFTN 1976
1486 SAKQLEKQNKQTEYELNVERELK--OALNGLSL-----TYSSSPSK--ROSL 1338
1487 SDRSRRLRELEIDALTOYDAEKANNOOKKNNKIKETLAKERYKPFSESEKTKREFLY 2036
1488 IDSLQOQVNSLQOOLADADROHEVIAVFTHL 1371
1489 VKLELDKRAKKADEBOORLTYENDLKRHL 2069

```

```

RESULT 5
MYH3_HUMAN STANDARD: PRT: 1940 AA.
AC P1055: 015492:
BT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 20-NOV-2001 (Rel. 10, Last annotation update)
DE MYOSIN HEAVY CHAIN, FETAL SKELETAL MUSCLE, EMBRYONIC
DE MYOSIN HEAVY CHAIN (SMHC).
GN MYH3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID:9606:
RX MEDLINE-69263803:
RX MEDLINE-69263803:
RA Rubinstein N.A., Kelly A.W., Sarkar S.
RA "Nucleotide sequence of full length human embryonic myosin heavy
chain cDNA."
RL Nucleic Acids Res. 17:3591-3592(1989).
RM SOURCE OF 774-1940 FROM N.A.
RX MEDLINE-90033298:
RX MEDLINE-90033298:
RA Elter M.S., Stehman H.H., Sylvester J.E., Ferteles S.H.,
RA Raychowdhury M.K., Rubinstein N.A., Kelly A.W., Sarkar S.
RA "Human embryonic myosin heavy chain cDNA. Interspersed sequence
conservation of the myosin rod, chromosomal locus and isoform

```

```

RT specific transcript of the gene."
RN FEBS Lett. 256:21-28(1989).
RN (3)
RP SEQUENCE OF 856-1940 FROM N.A.
RC TISSUE-Skeletal muscle;
RX MEDLINE-90235862:
RX Hober E., Buchberger-Steidl A., Braun T., Singh S., Goedde H.W.,
RX Arnold H.H.;
RT "Molecular cloning of three developmentally controlled isoforms of human
myosin heavy chain cDNA."
RN Eur. J. Biochem. 189:55-65(1990).
RN (4)
RP SEQUENCE OF 856-1940 FROM N.A.
RX MEDLINE-69366648:
RX Kersch-Hirachl I., Travis M., Blau H., Lelwand L.A.;
RT Expression and DNA sequence analysis of a human embryonic skeletal
muscle myosin heavy chain gene.
RX Nucleic Acids Res. 17:6167-6179(1989).
CC -1 SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (HMC) 2 ALKALI LIGHT CHAIN SUBUNITS (ALC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS OF THE MYOFIBRILS.
CC -1 DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
MUSCLE. THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT SEQUENCE INTERSPERSED WITH 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1 PPM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1 MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBDOMAINS (S1) AND 1 ROD-SHAPED
SUBDOMAIN (S2).
CC -1 SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 ROD-LIKE
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL consortium -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb.slb.ch).
-----
EMBL: X13988: CA33167.1;
EMBL: X13100: CA33492.1;
EMBL: X51593: CA335942.1;
EMBL: X15696: CA33731.1;
PIR: S04090: S04090.
HSP: P13538: 2MYS.
MIM: 160720:
InterPro: IPRO00048: IO;
InterPro: IPRO00049: Myosin_tail;
InterPro: IPRO01609: myosin_head.
Pfam: PF00052: IO: 1;
Pfam: PF00053: myosin_head: 1;
Pfam: PF01576: myosin_tail: 1;
PRINTS: PR00193: MYOSINHEAVY;
PRODOM: PD00035: myosin_head: 1;
SMART: SM00242: MSC1;
SMART: SM00243: MSC2;
SMART: SM00244: MSC3;
Kw: Myosin_muscle_protein;
Kw Calmodulin-binding; ATP-binding; Methylation; Alkylation;
Kw Multigene family;
Kw MYOSIN HEAD-LIKE;
FT DOMAIN 1 781
FT DOMAIN 782 811
FT DOMAIN 812 840
FT NP_BIND 840 1933
FT NP_BIND 119 186
FT DOMAIN 526 678
FT DOMAIN 728 758

```

```

FT MOD_RES 130 130 METYLATION (TBT-) (POTENTIAL).
FT MOD_RES 696 696 ALKYLATION (SH-1).
FT MOD_RES 706 706 ALKYLATION (SH-2).
FT CONFLICT 1331 1331 A -> G (IN REF. 3).
FT CONFLICT 1391 1392 KR -> QE (IN REF. 1 AND 2).
FT CONFLICT 1608 1609 SR -> RA (IN REF. 3).
FT CONFLICT 1663 1664 SR -> QT (IN REF. 3).
SQ SEQUENCE 1940 AA: 224035 MW: 43545864BAL253 CRC64;

Query Match 8.9% Score 623 DB 1 Length 1940:
Best local similarity 22.2% Pred. No. 1,6e-11:
Matches 304; Conserved 252; Mismatches 502; Indels 310; Gaps 40:

OY 241 TASNSNKGRLKKGKPSIQ-----QRLNSQMLDEYNTKSNQKHQIDOLEINE-- 291
DB 627 TPDQSGKRYKAKKSGSFQYNAFLRNLAKMSLRT-----THPFRKCLIPNETH 681
OY 292 -----DIKELRKTQ--BORILLNVNGLOLMEYVWADLESEFE 333
DB 682 FQAMHSYVLQRLNONGVLEGRICRCKPFRNLVDFQFQRYVLAASALLEGQFSRK 741
OY 334 KIKSLAKKQKHQESLRTITELKSNFRK-----YFSHSLQSGSFFKEDMLKQOQVY 387
DB 742 ACERLLASIDIDH-----TOYFQGHKVFYFAGLLGLEEM--DBRLAK-----L 785
OY 388 HTDSQSTQNPYHNSRSM-----TOYFQGHKVFYFAGLLGLEEM--DBRLAK-----L 785
DB 786 ITRQVNGRQFLMYERQKMGVQRESICQVNISSFMVHMHMKLFKIKPL--LK 842
OY 417 NQASVEBELLKKELEAMRFGDSAKORLQNLNLAHVACALALECERVKE--DS 473
DB 843 SAEERKAMTKKEEFQRTKDBLAKSEAKRELEKVLTVLQEKNDLOLOVQASESNLLA 902
OY 474 DEQIKOLEDALQVQKMTESQVYQOHTPLALCKHLTSDAQGNHRLMEELKQDLQ 533
DB 903 EFRQDL-----LAKQQLDKMLKEV-----TBRADEEETIAMELTA 939
OY 534 KRYVYQASAEVQKLAQIKQEMALVEEFKQDEGLKEMNKIKLOKLS-----MCELEER 588
DB 940 KKRRLD-----ECSELKRDIDLETLAVYERKEMTENYKMLLELSDGETIAKLT 996
OY 589 KRGKRLTEGQ-LQDLSA-----LALSTPAKFEKMSLSNELMEKAKKLDIVEREY 643
DB 997 KKA--LQAHQQLQDQDAEEDKYNKLSKTSKLOQYEDLESLSQEKIRAPVDLEN K 1053
OY 644 RSLNETRPLKRE-----LENLAKLQAOHKPRPE--OLSKRDEKQSG--ELCKRIET 694
DB 1054 KRLSQDLAKQESIDILENDKQDLKPKKDEYCOLQSKVEDBOYLGLOFQKRIEQ 1113
OY 695 SKNQTQLEIEKVCJDKRLKLTQGVNNLTTEKKNVPLVYS-----EMK-- 738
DB 1114 ATELELELEIABERTRAKTERKQNSQYARELELESERLQAGQVSTQTELEANKREAF 1173
OY 739 -----SHQYVDLKKKLSQ-----YHNYKTEKLEMKSL 769
DB 1174 KLRQDELEATLOHEAMVALTRKRIADQSVAELEQIDNLOQVNOKLEKESFELEIDS 1233
OY 770 MENASLSKVSRLVETPREPHEKQMA-----LKSNTLEKLOKJELNKKC--E 819
DB 1234 SSMSVYSKSNKLEKICRTLEQSLQSLANGKNEIRTSLSLTQSKRLQTEBAGELSQLE 1293
OY 820 DQETLTSLSM-----ENQDKTYSR-----QVYPT 847
DB 1294 EXEIVTSQSLSKQAFTQOTETRLQLEBENKAKNALAHLQSSRHCDLLRQV--EE 1350
OY 848 HEEIKTALSTLTKNRELVDYKRCQE-DINQEFKIKDENMLTKSLENTQNOVK--A 903
DB 1351 EDEKQAELOARLSKANSVQWKRYETDALQTELEBKKKKLAPRLOQESDOVEAVNA 1410
OY 904 EYISLSEHEKSG-----LKSKKVQDQNSAEILATYKKSQOEIVTLAHE 949
DB 1411 KQASLTKKQDQGEVDLMDVYBRANSTLAALDKQKQNTDVKVLAHKTKEESQAEHA 1470

OY 950 IAAQKRELTIOECIKLYA---PIISLECEKRFATKELKQLSQOTQVYTSDEEA 1006
DB 1471 SLKESRSLSL--ELFKMAVEAEALDQLETVRERNNYLEQIDQLQIANGKTHIE 1528
OY 1007 KCKGENDKKEELTLQDKQDKWHIENSYTEERASLRTEELRQKLDQKLYTEAK 1066
DB 1529 KSRQ-----TELEKADQLAEAEALAE--EAKIKRQLELTQVKSQEDDKIAED 1581
OY 1067 KKEKLEVEKMAKQTSLELAQOTLQKQVLPQVSEKSSQTSQTELEKELKRYCE 1126
DB 1582 EEELEL--KRYQORVEYH--QSLADVEARSRNEMALKKKMEQDNEIETQISLAR-- 1635
OY 1127 KEOQTVQLRQMLNENKNSVPLAEILOVEAFKQVGIKASLEKEEESQNTEDVK 1186
DB 1636 QAKETIKHLKLSVOOQLQDLOLADQLQDEGDELEQALV-----BRANLLQVBE 1688
OY 1187 LOSELQNTQALKXLETFREVDLYSYKA-----TQSLOTLSQIDMEKLEMANR 1235
DB 1689 LRATLEQTERA--RKLAEQELDSNRVQLLHTQNTSLHTFKKLETFQDMLQDSVEDAR 1747
OY 1236 KYEECEVEYLA-----KKKELESAKDEKLHPSIQEIKDOQERCKS----- 1279
DB 1748 DARRKAEKAKATIDAMMAEELKEEDQTSAMLEKMKKLQVQDLOHLELBOQLK 1807
OY 1280 --LTTTELQRTQESQKQJANDKLTIELMDYERLQKALNSQSLQYSSSK--R 1334
DB 1808 GKGKQIKLETRHLELEFELGQKNT--SVKIGRYERVRVELTYQSEDERKMYLR 1864
OY 1335 OSQQLDSILOQVRSLOQQLADARQHOEVATAYR--THLSLAQGHMD 1380
DB 1865 LQDLVDKLOVYKYSKQKQAEDEQDANHLTFKRAQHELEAEERAD 1912

RESULT 5
USOL_YEAST
ID USOL_YEAST STANDARD: PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INTRACELLULAR TRANSPORT PROTEIN USOL.
CN Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OC NCBI_TaxID=4932;
RN 11
RN SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RC MEDLINE=91185402; PubMed=2010462;
RA Nakajima M., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RT J. Cell Biol. 113:245-260(1991).
RN 121
RN SEQUENCE OF 782-1790 FROM N.A.
RA Kosterlitz M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
RA Kennedy K.E.; to the EMBL/Genbank/DBD databases.
RN Submitted (Feb-1993)
RN SOURCE OF 17-93 FROM N.A.
RA Bel Y., Symington L.S.;
RA Submitted (May-1996) to the EMBL/Genbank/DBD databases.
CC - FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC - SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF DOMAINS THAT REPEAT. REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.

```

```

CC -1- SIMILARITY: BELONGS TO THE VDP/USOI/YBL047C FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use. The copyright for this entry is held by the EMBL outstation. The
CC modified and this statement is only as long as its content is in no way
CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54378; CAA38253.1;
DR EMBL: U03188; AAB00143.1;
DR EMBL: U03668; AAB6659.1;
DR PIR: A38455; A38455.
DR SSI: S50022.1; S501.
DR SGD: S00022.1; S501.
DR Interprot: IPH002017; Spectrin.
KM Transport: cyto. stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724
FT DOMAIN 725 1790
FT DOMAIN 465 487
FT DOMAIN 991 1790
FT DOMAIN 1172 1786
FT DOMAIN 924 894
FT CONFLICT 1253 1255
FT CONFLICT 1319 1319
FT CONFLICT 1461 1461
FT CONFLICT 1581 1581
FT CONFLICT 1600 1600
FT CONFLICT 1661 1661
FT CONFLICT 1772 1772
FT SEQUENCE 1790 AA: 206424 MW: 662B21659FD4818 CRC64:
Query Match 8.88; Score 618; DB 1; Length 1790;
Best Local Similarity 22.18; Pred. No. 2,1e-11;
Matches 278; Conservative 257; Mismatches 461; Indels 260; Gaps 43;
OY 219 LGHDS-----SYARIQDMLD-ILT-----LAKTSNSKMGK 250
DB 653 LGDNVNASRIKQFKDYSFVKDMESILFPELDGLPKVYFTFOLDKENTIRIR 712
OY 251 ELMKGSQ-----QANLSMDLVNTKSNQK--NONTD-----LEIENE 291
DB 713 TALSHDDEPIKISFEVEKILQOCTIKETISLOTETSHNTETKILATINERK 772
OY 292 DLAEFLKRIQOORILILDKVNGIOLNREVVADOLESEKELSLAAKKEHPSIR 351
DB 773 ELDERYQIOLNSSHSSLKEMFSLIETELK--VDSID--EMTQIOLVETKMKENGTAL- 827
OY 352 TIEALKSRKFEYSIDHLSGSHRKEDMLIKOGQHTWDSCTGCPVHMOSKMLAPL 411
DB 828 -----LEYKSTIH-----KQDSI-----KTLDEKOL 848
OY 412 BLALPNQASYSNEI-----LKELEAARFFCDASAKODULK-----LONE 451
DB 849 ETLT-SQKKAKEDQINKKQKDFALSBEMQAVENKSKEDKDSVNHOKETSKJED 907
OY 452 LAHVACCAALAECEKRYEVEDSQIQIOLDALQVQKRYESEKYNQOCHTFLAKHE 511
DB 908 IAKTITELKINENLEBKITCOHNSKNEHISHELVE--TKS-----FROSH-----DN 955
OY 512 LTSDAAGNRLMEELK---DQCKMYKKE-----AAAEKGLANDIONOM 557
DB 956 LVA-----KLTEKLKSLANNKDMQAKESLAKAVESSNESSIOLSNOKLDSMO 1008
OY 558 LVEERKDRBKIMEKRRLOELSMCELE-----FRCKRIEMEGQOIKSLAKLALSP 613
DB 1009 EKNQVLEKSEKLENTKQITSLQDTKEEITISQSSKQDYSOISLKEKLELAT 1068
OY 614 AKFKNKSLSLNEMKAKKLLDVEERPELSLWTRPKLEKELKAKKLAQHYKKEP- 672

```

```

DB 1069 AND-ENV-----NRSELKTRSELEALAAVKKLAKNELFKLETSKALKEVENEHL 1122
OY 673 EQLSRLEKSGEGLERITELTSKQNTLOKEIEYVCLDMLKLTQOVNNLTBMKNPLKV 732
DB 1123 KEKTIQLEKATEKQOULMSLRANLESELEKEDHQAOLKKEVQIDANKERO----- 1174
OY 733 SEKKKSHDIVDOLAKKLSVTHKTYETKELLMKELLMENASLKNVSRLETVPIPERH 792
DB 1175 -----YNEISQIOLDELITSIQOENSLIKKNDLSEBEKYMANKST-----S 1214
OY 793 EREMAKLKSNITELKQKLSLEKCKGEQ-----EKIYSJMEENMLCKTTHSOVVPKTH 848
DB 1215 EEOULMKSEKIDALANLOIKELKAKENNEENSLISIKSVSEYIVKLEIODECCNFKEKV 1274
OY 849 EETKALSTLQKTNELVADVKKCEQDINOEFVKIDENELKRLNLENTONOVAAEYISL 908
DB 1275 SELEKTLAKSEDK--NSKLTQLOESEKIKE-----DAKTELTLQLETKINLSAK-- 1326
OY 909 REHEKSGSLPKSKMKKQVDMNSAEILAYKSGOEYVTLHEPAAKOEIDTIOGCIKIKY 968
DB 1327 EKESELSIRKTSSEKRNFAEDQLETK--NEIQIKQAFKRRKILBESSITOEY 1383
OY 969 APTISLEDCERKFKATEKLEKEDLSQOTQVYFSEEEKAKCKOENDLAKE-----LTLQ 1024
DB 1384 SEXINTLEDELRLQVENEK-----AKEDINTSEKVSLSNDELKEKQVNTKISLO 1437
OY 1025 KDI---KDK--NVHIEKSVFETRLSKRTLELNRQLOLQNTTEKKEKELVBNQOT 1080
DB 1438 DELSYQDKITNDEKTLSTIRBNRQDLSLKEQLMAQESKAVYEDQKTLKESSERK 1497
OY 1081 SETLAQOTLLOKQHVPLQDQVESLKSLSGTIEPILEKELQKCYERDQVITOLQMLE 1140
DB 1498 AE-----LEKSEKMKKLESTIESMTLELS-----SMEITIRSDKLE 1536
OY 1141 NQNSSVPLAEHLYQKAEPEKGVITASLRKE--EESQNTKEVSKLOSIQNTQALK 1199
DB 1537 QSKAKAEEDIKILO--HEKSOULSRKNSKQIDELASLALIAKSSGSELEVQKBLN 1592
OY 1200 KLETRVVDLSKRYKATRSQLETOISDNLKELNANRKREYECVETLAKKKELEANDRE 1259
DB 1593 NAOEKIRINAEENYVLKSKLEDEIERELKQAKAEIKSNQEE--KELTISRLEKEL----- 1643
OY 1260 LHFISIDIEKIDQOERCQSLTITIELQRRIOESANQILEAK-----DNKTIELADVBRK 1315
DB 1644 -----DELDSTQOAKQNSSEERBAERFQVSKQDUEKMLALETKRYVADKQKMD 1697
OY 1316 QALNLTSLQVLOSQSPSKRSQSLDISLQOQVNSQLOLADQMLADQVNAVIRPH 1371
DB 1698 RBDYVAKKTTDSQROELTEKELAKELDNKAKNSKIKE--ANEDRSSIDQMLLVYDL 1751

```

```

RESULT 7
MYTH_HUMAN STANDARD. PRT; 1935 AA.
ID P12581; O14904; Q08373.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM (MYHC-BETA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominoidea; Homo.
NC C12410-3606;
LN 11;
RX SEQUENCE FROM N.A.
RX MEDLINE-91065634; PubMed-2429844;
RX Jaenike T., Dieleick K.W., Haas W., Schleich J., Lichter P.,
RX Pfordt M., Bach A., Vosberg H.P.;
RX "The complete sequence of the human beta-myosin heavy chain gene and
RX a comparative analysis of its product.";
RX Genomics 6:194-206(1990).

```

RP SEQUENCE FROM N.A. PubMed-2362820;  
 RA MEDLINE-90301496; Yamuchi-Takihara K., Kellam B.,  
 RA Llew C.-C., Sole M.J., Llew J., Llew J.,  
 RA Anderson D.H., Lin L., Llew J.,  
 RT "Complete sequence and organization of the human cardiac beta-myosin  
 RT heavy chain gene.";  
 RT Nucleic Acids Res. 18:3647-3651(1990).  
 RA SEQUENCE OF 1-115 FROM N.A.  
 RA MEDLINE-89264452; PubMed-2726733;  
 RA Yamuchi-Takihara K., Sole M.J., Llew J., Ing D., Llew C.-C.,  
 RT "Characterization of human cardiac myosin heavy chain genes.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).  
 RA [14]  
 RA ERKTRUM.  
 RA Tomuchi-Takihara K., Sole M.J., Llew J., Ing D., Llew C.-C.,  
 RA Llew C.-C., Sole M.J., Llew J., Llew J.,  
 RA [15]  
 RA MEDLINE-86176778; PubMed-2421254;  
 RA SEQUENCE OF 1310-1935 FROM N.A.  
 RA MEDLINE-86176778; PubMed-2421254;  
 RA Szei L., Lelwand L.A.,  
 RT "Characterization of diverse forms of myosin heavy chain expressed in  
 RT adult human skeletal muscle.";  
 RT Nucleic Acids Res. 14:2951-2969(1986).  
 RA [16]  
 RA Lelwand L.A.,  
 RT "Myosin heavy chain genes.";  
 RT Submitted (MAR-1998) to the EMBL/Genbank/DBD databases.  
 RA [17]  
 RA SEQUENCE OF 1410-1935 FROM N.A.  
 RA MEDLINE-88299163; PubMed-2969939;  
 RA Kuriyabashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.,  
 RT "Molecular cloning and characterization of human cardiac alpha- and  
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of  
 RT their expression during development and pressure overload in human  
 RT atrium.";  
 RT J. Clin. Invest. 82:524-531(1989).  
 RA [18]  
 RA SEQUENCE OF 785-1935 FROM N.A.  
 RA TISSUE-Skeletal muscle;  
 RA MEDLINE-9033582; PubMed-1691900;  
 RA Robber E., Buchdorfer-Geid A., Braun T., Singh S., Goedde H.W.,  
 RT "Cloning and characterization of a human ventricular cDNA library and  
 RT characterization of a beta myosin heavy chain cDNA clone.";  
 RT J. Gene. 76:47-53(1987).  
 RA [19]  
 RA SEQUENCE OF 1393-1935 FROM N.A.  
 RA MEDLINE-87192738; PubMed-3032769;  
 RA Jandreski M.A., Llew C.-C.,  
 RT "Construction of a human ventricular cDNA library and  
 RT characterization of a beta myosin heavy chain cDNA clone.";  
 RT J. Gene. 76:47-53(1987).  
 RA [20]  
 RA REVIEW ON VARIANTS.  
 RA MEDLINE-9603076; PubMed-8533830;  
 RA Arai S., Matsuo K., Hirayama K., Sukurai H., Tamura M., Ozawa T.,  
 RA Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawano M., Takao A.,  
 RA Hosoda S., Komma K.,  
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in  
 RT hypertrophic cardiomyopathy.";  
 RT Am. J. Med. Genet. 58:267-276(1995).  
 RA [21]  
 RA VARIANTS CMI-GLN-256 AND ARG-741.  
 RA MEDLINE-93248216; PubMed-8480925;  
 RA Pennapazari L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.,  
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central  
 RT core disease in hypertrophic cardiomyopathy.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).  
 RA [12]  
 RA VARIANT CMI-GLN-403.  
 RA MEDLINE-95179132; PubMed-1975517;  
 RA Goldfarber-Lawrence A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,  
 RA McKenna M., Seidman C.F., Seidman J.G.,  
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta  
 RT cardiac myosin heavy chain gene missense mutation.";  
 RT Cell 62:999-1006(1990).  
 RA [13]  
 RA VARIANTS CMI-GLN-249; GLN-403; CYS-453 AND MET-606.  
 RA MEDLINE-92201193; PubMed-1552912;  
 RA Watkins H., Rosenzweig A., Heang D.S., Levi T., McKenna M.,  
 RT "Characteristics and prognostic implications of myosin missense  
 RT mutations in familial hypertrophic cardiomyopathy.";  
 RT New Engl. J. Med. 326:1108-1114(1992).  
 RA [14]  
 RA VARIANTS CMI-GLN-403; CYS-453; ARG-584 AND MET-606.  
 RA MEDLINE-94070853; PubMed-8250038;  
 RA Watkins H., Thierfelder L., Aman R., Jarcho J., Matsumori A.,  
 RA Seidman J.G., Seidman C.F., McKenna M.,  
 RT "Independent origin of identical beta-cardiac myosin heavy-chain  
 RT mutations in hypertrophic cardiomyopathy.";  
 RT Am. J. Hum. Genet. 53:1180-1185(1993).  
 RA [15]  
 RA VARIANTS CMI-GLN-403 AND VAL-908.  
 RA MEDLINE-92346810; PubMed-1638703;  
 RA Epstein N.D., Cohn G.M., Cyran F., Pennapazari L.,  
 RT "Differences in clinical expression of hypertrophic cardiomyopathy  
 RT associated with two distinct mutations in the beta-myosin heavy chain  
 RT gene: a missense mutation and a 403Arg-->Gln mutation.";  
 RT Circulation 86:345-352(1992).  
 RA [16]  
 RA VARIANTS CMI-LEU-403 AND TRP-403.  
 RA MEDLINE-94075629; PubMed-8254035;  
 RA Dussan E., Komajda M., Feiler L., Dubourg O., Dufour C., Carrier L.,  
 RA Vysnky C., Becovic J., Hengstenberg C., Al-Mandawi S.,  
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and  
 RT identification of a hot spot for mutations in the beta-myosin heavy  
 RT chain gene.";  
 RT J. Clin. Invest. 92:2807-2813(1993).  
 RA [17]  
 RA VARIANT CMI TRP-403.  
 RA MEDLINE-94093568; PubMed-8268932;  
 RA Koolman J.C., Brink P.A., Corfield V.A.,  
 RT "Identification of a new missense mutation at Arg403, a CPG mutation  
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in  
 RT hypertrophic cardiomyopathy.";  
 RT Hum. Mol. Genet. 2:1731-1732(1993).  
 RA [18]  
 RA VARIANT CMI ASN-615.  
 RA MEDLINE-93036688; PubMed-1417858;  
 RA Nishi H., Kimura A., Harada H., Sasazuki T.,  
 RT "Novel missense mutation in cardiac beta myosin heavy chain gene  
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";  
 RT Biochem. Biophys. Res. Commun. 188:379-387(1992).  
 RA [19]  
 RA VARIANT CMI GLY-778.  
 RA MEDLINE-93343938; PubMed-8343162;  
 RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.,  
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked  
 RT to familial hypertrophic cardiomyopathy in affected Japanese  
 RT families.";  
 RT Biochem. Biophys. Res. Commun. 194:791-798(1993).  
 RA [20]  
 RA VARIANT CMI VAL-908.  
 RA MEDLINE-93164853; PubMed-8435239;  
 RA Nishi H., Kimura A., Toshima H., Sasazuki T., Miyamopoulos P.,  
 RA Gilligan D., French J., Choudhury G., Williamson R., Oakley C.,  
 RT "Identification of a mutation in the beta cardiac myosin heavy chain  
 RT gene in a family with hypertrophic cardiomyopathy.";  
 RT Br. Heart J. 69:136-141(1993).  
 RA [21]  
 RA VARIANT CMI TRP-719.  
 RA MEDLINE-95179132; PubMed-1974131;  
 RA Greve G., Bachinski L., Friedman D.L., Czernezewicz G., Aman R.,  
 RA Towbin J., Seidman C.F., Roberts R.,











|             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |           |                                                              |                        |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------|------------------------|
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 946       | EIAAKRRLDLDGCIKIKATPAITSLSECEPKRKFKATELKASQOLSGOOTOXNTPSEAAX | 1008                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1411      | SROEKNASTLST-ELFKAFAVEESLEHLE-TREKKNNKLDEIISDLTBOLOGCKAYHE   | 1527                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1009      | CROENKRALKETLTQNLIKNDYNIENSYSFKALSKTEELRNOLDLOKYVAAKE        | 1068                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1528      | LKRIKQLIEVKLEILQSILEVASLSHE--GGAIIMAPLFNQIKAEIRLAODE         | 1584                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1069      | KEXLYEANNKTOSSELLAQTLLAQONWPLDYEVASKSGLTGITEFLKEELTKORCTEK   | 1128                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1585      | MEOAKRMHLRWVSCL-QTSIDAFETRSMNALRYKKMKSCULMBEHFILOSANMISA     | 1641                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1129      | OOTVOYLROMLENOKNSNVYLEMYOFARFEVGIIKKS---LRKEEPSO---NKPE      | 1182                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1642      | QR--HKKNSQAHLKDPTOLOLDGVANDJLKNAIVERNINMLDAELEVAEDTES        | 1658                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1183      | VSYFL-OSETIOHTALQIALKETREVVDISKYATRSD--LETQISDNLERLANIKRYE   | 1238                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1659      | BKRLMDLPETSERADQNTSLTNOKKMSBSIDLTDLEVAVOCNRMEERKA            | 1758                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1239      | E-VCEEVLNL-KKEELAAMDKELPITSIPKOHOECRS-----LTTIELD            | 1287                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1759      | KAITDMAAMELEKIETDSOTSAHLERRKMIMEOTDIQQIHDLDEAGAILKGKODALE    | 1818                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1288      | RPIESAQIDEADKRTTELTLWDYKERKALOMALSQUTYGSSGSJK---ROSOLDISIQQ  | 1344                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1819      | AHVLELELDELAKORKNE-SYKCMRRSRKITELTYTEDKRNLMRLDTLVKIDLI       | 1875                   |
| Oy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1345      | GYSBISOULDADRNOGHVTAIR-THLSAAGOCHMD                          | 1380                   |
| Dy          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 1876      | KYYAKYRAOEAREEQANTNMSPFRKHGEHDLAEGRBD                        | 1913                   |
| <hr/>       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |           |                                                              |                        |
| RESULT 11   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |           |                                                              |                        |
| ID          | MYH9_RAT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | STANDARD: | PRT,                                                         | 1961 AA.               |
| CC          | CG68112                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | (Ref.)    | 40                                                           | (Created)              |
| CC          | 20-AUG-2001                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | (Ref.)    | 40                                                           | (Sequence update)      |
| CC          | 20-AUG-2001                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | (Ref.)    | 40                                                           | Last annotation update |
| CC          | 20-AUG-2001                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | (Ref.)    | 40                                                           | Last annotation update |
| CC          | PROSIN HEAVY CHAIN, NONMUSCLE TYPE A [CELLULAR MYOSIN HEAVY CHAIN,<br>TYPE A] (NONMUSCLE MYOSIN HEAVY CHAIN-A) (NMHC-A).                                                                                                                                                                                                                                                                                                                                                                                |           |                                                              |                        |
| GN          | MYH9-                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |           |                                                              |                        |
| OC          | Rattus norvegicus (rat).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |           |                                                              |                        |
| OC          | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;                                                                                                                                                                                                                                                                                                                                                                                                                                       |           |                                                              |                        |
| OC          | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.                                                                                                                                                                                                                                                                                                                                                                                                                                   |           |                                                              |                        |
| NCBI_TextId | -10116;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |           |                                                              |                        |
| LN          | [1]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |           |                                                              |                        |
| PR          | SEQUENCE FROM N.A.C. Tsch K., Adelstein B.S., Benven M.A.;<br>Choi O.H., Ray J.D.<br>Submitted (Jul-1995), to the EMBL/Genbank/DDBJ database.                                                                                                                                                                                                                                                                                                                                                           |           |                                                              |                        |
| RU          | -- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,<br>CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND<br>CAPING.                                                                                                                                                                                                                                                                                                                                                           |           |                                                              |                        |
| CC          | -1 SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY<br>CHAIN SUBUNITS (MHC). 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2<br>REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).                                                                                                                                                                                                                                                                                                                       |           |                                                              |                        |
| CC          | -1 DOMAIN: THE RODLIKE THAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING<br>SIMILARITY WITH A 28 RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,<br>CHARACTERISTICALLY UNPHASED BY TWO FULLY CONTROLLED CDS.                                                                                                                                                                                                                                                                                             |           |                                                              |                        |
| CC          | -1 SIMILARITY CONTAINS MYOSIN-GLOBULAR HEAD DOMAIN.                                                                                                                                                                                                                                                                                                                                                                                                                                                     |           |                                                              |                        |
| CC          | -1 SIMILARITY CONTAINS 1 TO DOMAIN                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |           |                                                              |                        |
| CC          | This SWISS-Prot entry is copyright. It is produced through a collaboration<br>between the Swiss Institute of Bioinformatics and the EMBL Outstation -<br>the European Bioinformatics Institute. There are no restrictions on its<br>use by non-profit institutions as long as its content is in no way<br>modified and this statement is not removed. Usage by and for commercial<br>entities requires a license agreement (See http://www.isb-sib.ch/announce/<br>license.html or license@isb.sib.ch). |           |                                                              |                        |
| EMBL:       | U31463; AM274950.1 ;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |           |                                                              |                        |

DR HSP: P10587, 1981.  
 DR InterPro: IPR000048: 10.  
 DR InterPro: IPR002928: myosin\_tail.  
 DR Pfam: PF00612: 10: 1.  
 DR Pfam: PF00063: myosin\_head: 1.  
 DR Pfam: PF01576: myosin\_tail: 1.  
 DR PRINTS: PR00199: MYOSINHEAVY.  
 DR PRODOM: PD001355: myosin\_head: 1.  
 DR SMART: SM00242: MYSC1.  
 DR PROSITE: PS50096: 10: 1.  
 DR Myosin: ATP-binding; Calmodulin-binding; Actin-binding;  
 Colled coil; Alkylation; Multigene family.  
 FT DOMAIN 1 778 808  
 FT DOMAIN 1 808 808  
 FT DOMAIN 1 808 808  
 FT DOMAIN 1 808 808  
 FT MOD\_RES 654 676  
 FT MOD\_RES 694 694  
 FT MOD\_RES 704 704  
 SEQUENCE 1961 AA: 226336 MW: 998760968199 CRO64;

Query Match 8.6%; Score 606; DB 1; Length 1961;  
 Best Local Similarity 20.8%; Pred. No. 5.1e-11;  
 Matches 271; Conservative 271; Mismatches 478; Indels 284; Gaps 48;  
 189 TALLMCCGCGCD--AVEVLKNDADYVLDLADGDSYVIRGIDNLDITLTKASENS 246  
 780 TDVITGFACCRCGLYAKFAKFRQDOLTKNKKVLDQNCAMVRLT-----RW 825  
 247 NKQELMKKSPSLQORNLQOMLDENVTKSQNQHONITDLETENEDLKRYKTOEOR 306  
 826 QMRLKLTAVRL--NSTRIEDLAKMELTKYREKILALEN-----RLTEMQSO 877  
 307 LLDKVGSLQLEMYVADLSEKFKSKILAKKQKHESLRITLAL--KSPFY 362  
 878 LMAE--KLDQDQDQATLCEALAEVRLARGLTAKKQELFETQIDRLAVREEREKCY 933  
 363 RESDLSGSHFKEDMLKQGMVYTSQSTSTQMPVH-MQSKMLRPLRLPQASV 421  
 934 IQAK-----KKMQQNTDLEQDLEESAROKTDLKRYTALMKLE----- 977  
 422 SPNTLKEELMARTPCDSADQRL--KQMLKALVAKQKALASCVNEDSE 475  
 978 -EDQIIMEDQN--CKLAKKKLLERVAEFTDLMEEBKSKLA--KLNKHA 1027  
 476 QIKQDLADLVQKRYESEGKVNQKQTHFALKEHTSDAQTNRIMELKIDOLKDK 535  
 1028 MTLDEBYLRREKROBELKTRKLEEDSTDSQIA-----ELQDLK 1075  
 536 V--VYGSASANGKLRQIKQWYVPRKQDGLKQKELKSLKSLERKSG 591  
 1076 MQLAKKEELQMLARVEEAKQKNAKLRLELTQISLEQDLESSRA--CNKAKOK 1134  
 592 RL-----TEMQQLKDLAKLALISPAKEFNNKSLSLKLEAKK----- 633  
 1135 KLDCEELALNKTLEDTLDSTVAQOEL--RSKRQDQVSLKTLKDYKDAHQEOENR 1191  
 634 KLDVKEKESKLNELRPL-----KRELNKALALQNYK-----PEHEOLK-- 676  
 1192 QKNSQVAYELADQLQKRYKATLEAKQCLINERQALANVYKALQKQSGSHKRRKE 1251  
 677 ---SRLEQSGELGKRI--TELSSNQTKQETKCYCDKDKLLTQ--QVNNLTPEMKNP 729  
 1352 AQDLQVQKFSF--GERVTELDADVSKLQVLEDSV--TGLLNOSKSKSLKQFSS-- 1304  
 730 LKVEEMKSHDVTVDLKKLS--DYTHKYTEKLEMKLIMENASLKNVSRLETVF 786  
 1305 -ALSDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 1356  
 787 IPRRIKEMKALNSITTELKQSLSEKQKSGDEQNT--YSLKEMNDLKKTKMSQVYV 845

DB 1357 -----EKQATLAAQVDTMKKMKED--GVGCELTAEAEKRRHLQKDLBGLSQRLEKRYAAV 1409  
 846 KTHKEETKALSTLDTNRELIVYKKEEDIN-----QEEVAKIDENETLKRNLNTON 939  
 1410 DKLEKTKTRLQOEDLDL--LVLDHQRQSVSNKERRKQKFDQLLAEKTKIAKVAEED 1466  
 900 QVAKETSLPERHEKSGSLKSKKAYQDMSA--TLAKTKQDEIVTLHEAIAQRE 956  
 1467 RAALF--ANEKFKLSIALAEAEQKALERLRLNQFRTFEMDLMSSKDQVGSVHE 1523  
 957 LDITQECIKTKTAPV--ISLECEKRFKATP-----KELKSQLSQOQTKYNSSEE 1005  
 1524 LKESRRLADQVEEMKQTLDELDELQATDQKLRLEVYLQAKKQAFQEPLOGSDQSE 1983  
 1006 AKK-----CKQDNKIKKEL--TLODKQKVVHINSETERALSKEE 1050  
 1584 KKKQOLYRVEMENAEDEBRQNSIANKAKKLEMDQLEMDLQATKAREALIQOLR 1643  
 1051 LKQOLKDLQNTAKKKEKLEVEANQKTSRSEIILAQTLQKQHPVLEQVESLKSLSGT 1110  
 1644 LQAKQDQMDHDVDTASREBELAQ--ANEKREKLL-----KSNRE 1682  
 1111 ETLKEELTKQRCYKKEQDQVTVQLOKLENNKSVPLAHQVNEAFKEVQIIMASL 1170  
 1683 MIDQDEBLAAKRAKKNQAO--EDELQDETASNSKQKALAEKRRLEALMIDBEL 1739  
 1171 RKE-----PESQNTREVSQKQSEI QNTRQALKKLETR 1204  
 1740 EESQNTLJNRLKAKNQLQIDQINTDLNLSHQKQENMAKQOLEQN--KELAKLDQM 1798  
 1205 EYVDSLTKRATQSLDETQISDLNEKLANLRKYEECEVLIARKK--ELSKDKBEL 1260  
 1799 ESNVSKTKASTALENKTLQDLEQDQNERQKQAKQVNRKAKLLQVLYQDSEKQ 1857  
 1261 LHSFIEDQITDQORCKQSLTTELQORLOS--AKQIDANKITLIDLYNDERLQKA 1317  
 1858 ---NADQPDQ--ADKASRLQKLRQLEAEFEQRQVANAARRLQRELEDA----- 1904  
 1318 LNLGSLQTYGSGSPSKQSQLDSIQDQVNSIQDQDLAD 1357  
 1905 -----TETADAKREVSLSLKKLRQD 1926

RESULT 12  
 MY9\_HUMAN STANDARD; PRT; 1960 AA.  
 ID MY9\_HUMAN STANDARD; PRT; 1960 AA.  
 AC P35579; 060805;  
 DT 01-JUN-1994 (rel. 29, Created)  
 DT 15-DEC-1998 (rel. 37, Last sequence update)  
 DT 20-AUG-2001 (rel. 40, Last annotation update)  
 OS Homo sapiens (Human).  
 GN MY9.  
 GN TYPE 1 (NONMUSCLE MYOSIN HEAVY CHAIN-A) (NM\_001-A).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RP SOURCE: PubMed=10591209;  
 RA Boudreau-Morissette, R., Collins J.E.,  
 RA Champ M., Smink L.J., Ainscough R., Almeida J.P., Babarge A.,  
 RA Baguelin S., Bailey J., Barlow K., Bates K.N., Beasley O., Bld C.P.,  
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Buttrill W.D.,  
 RA Burton C., Carter N.P., Chen Y., Clark G., Clegg S.M.,  
 RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corry N.,  
 RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhimi P.D., Dockree C.,  
 RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,  
 RA Fleming K. French L., Garner A.A., Gilbert J.G.R., Goward H.B.,  
 RA Graham D., Griffiths M.N., Hall C., Hall R., Hall T.M.,  
 RA Hechtolt R.M., Ho S., Holmes S., Hunt C., Hunt J., Knappe J.,  
 RA Kimberley A., King A., Laitin G.K., Langford C.F., Levenshtein M.A.,







|    |      |                                                               |        |
|----|------|---------------------------------------------------------------|--------|
| Oy | 398  | MYTINOSHMLRLELALPNASHSITNELLAKKEEMARTPOSAKODKLONELAKRYA       | 4357   |
| Dp | 931  | -----EEDENENEL-----                                           | YA 944 |
| Oy | 458  | ESKALALDECVENEDBOEIOQLDOLADVOYKRYASBGVOYKOHOTFLAKKEHLSIDAA    | 5175   |
| Dp | 943  | KRYLDEKSSKLKRIDID-----LELTLYKKEKNTENKYSSEEMALENISK--L--       | 996    |
| Oy | 518  | TOHRLAMELOO--LAKRYVYBDSAPVKRLMOIKONELVIEFKROBKLMEKRL          | 576    |
| Dp | 997  | TEKKLODAMQOTLIDLOY--BEKONKIKINKALDOODLODLOSL--BOEKKI 104      |        |
| Oy | 577  | OKEV-----SMELERPK--GKRLTEKLODLODLOSLALISPAKRE 615             |        |
| Dp | 1049 | RADLEAKRLEBDLAKSOESIDONLEENKOOTEKLLKKEPELSLOAI-----DDE 110    |        |
| Oy | 619  | NNKSLLSNEKAKLILVEREYSALETAPLAKKEYLKAKLAOYKPEEHQOKSR 676       |        |
| Dp | 1102 | OYHSL--OYOKITELONLELEBEZIDELNTIKALIKOSHOLAN-----ELEISER 115   |        |
| Oy | 679  | LKQSSLOEGRITELYSK-----OTLOKEPELOLOML-----LOYKY 719            |        |
| Dp | 1154 | LEKSAVSAQOI--ENKKKEAEFOKMRRLDELEATLOHAPATLAKKRODSVALEBOD 122  |        |
| Oy | 720  | NITTEKRWPLKYSBEKSSIDVY--VODANKLSO-----YTHKYEEK 767            |        |
| Dp | 1213 | NLOQYKAKERKESKALIDMANSALSSKSIETRCROKMAOSPTLAKKEDQOT 126       |        |
| Oy | 763  | -----LEMK--LIMNALSIMSLFETVTPPRERKEKMAOSNTIKOTLOSELAK 816      |        |
| Dp | 1273 | LILHLOMKARLOQONELSHRYEKESLI--SOLTSKROLTOOELTROMEKELK-- 132    |        |
| Oy | 817  | GEDOEKITYLSMBNDKLAKKMSHOYUWKTEHITVLSLTOKTMBELDYKKE-D 875      |        |
| Dp | 1339 | -----KANNALALOSBHC--DLEBOY--BEBOALALORALSNANSEVOKKRTETD 138   |        |
| Oy | 876  | INDEPKAKDEMLKRLKRLNTOYU--ANYSILIEHEKMSG-----LAKS-- 921        |        |
| Dp | 1383 | ALOTHELEAPAKKILORLOAEFEETENKASCSLEKTOQLOEYDLMBLOLBSHRA 144    |        |
| Oy | 922  | -----KKVODMSALIKAKKSOEIVYLIEBELAOKRYELOETOKL--VAPITSLER 976   |        |
| Dp | 1443 | CATLDKROKFOKRYALAKOKLOE--SOELEAPAKOESHSLETPRYMNAVEYUOLE 103   |        |
| Oy | 977  | CERFKATFEELKLOSOOTQNTNSEENKAKOENDKLLELLOKDKOYNIEN 105         |        |
| Dp | 1501 | TILR-----ENKQDELSIDLOFQ-----AELONKALAKKLLEKDEKSOALALEB 155    |        |
| Oy | 1037 | SEYTERA--LSRTEBELSRLDLOKOTYAKKEPELVENSKOTSELSLAOFTLOLOY 109   |        |
| Dp | 1554 | SLIEBESLILMLOYSKSELEDKRYEKELEIBOL--KNSOANAE--ALOSVLEAIR 161   |        |
| Oy | 1096 | PLADYSSKSLISGTELEKELATKROCEBOOTVTOLOKLEONKSSVLAIBOY 115       |        |
| Dp | 1611 | SHNDALRKKKEDLOMBEYLOGHSNR--QANQONHILFTVOODKSOILDLMDLMS 166    |        |
| Oy | 1156 | KPAEEYGVITASLARKKEESONTPEKSKOSELONTOMAKLLEREYDLS----- 121     |        |
| Dp | 1668 | NEDAKSOLAV-----ERNNDLLELEELMKYALBOYTER--FRISDELLDASDYOL 171   |        |
| Oy | 1211 | -----KTKATSOLOETLOENLKNY--NKRIEVECEYL-----NANKKEESA 125       |        |
| Dp | 1720 | LHSQNTLITKTKLNDLOHQAQVAVENSIQOESNANKKATLADMAAMELEKQOT 137     |        |
| Oy | 1355 | KOBEELSLFIOEIKDOOPEKSKSLTTTELEPRIOKSOETAKDKITELML--DYE 131    |        |
| Dp | 1780 | SAHLEPKMLLOTVYDLOHLEDA-----EOLAKGOKIOKLENNRVELIEHLEDA 183     |        |
| Oy | 1313 | PLK--QALNC-----LSOYTGSSGSPK--ROSOILDSLOOVRSLOOVLADROY 1359    |        |
| Dp | 1834 | OKROBALKAKKERTKENTIKTONOEDBKNIILALODLOVLOKAKKSKIKNOBEBED 1892 |        |





Thu Apr 11 07:43:36 2002

us-09-750-590-2.std.rsp

Page 22

Search completed: April 10, 2002, 18:11:32  
Job time: 165 sec

[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: April 10, 2002, 18:09:52 ; Search time 64.08 seconds

With 13197.995 Million cell updates/sec

Title: US-09-750-590-2

Perfect score: 7039

Sequence: 1 MMSCFGAPKRRQADWKNK.....DYQALLQIIHQKGLVCSA 1401

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: SP:archaea:.\*  
2: SP:bacteria:.\*  
3: SP:fungi:.\*  
4: SP:plantae:.\*  
5: SP:metazoa:.\*  
6: SP:invertebrate:.\*  
7: SP:chordata:.\*  
8: SP:vertebrate:.\*  
9: SP:phage:.\*  
10: SP:plant:.\*  
11: SP:rodent:.\*  
12: SP:virus:.\*  
13: SP:vertebrate:.\*  
14: SP:unclassified:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | Score  | Query Match | Length | DB ID    | Description        |
|--------|--------|-------------|--------|----------|--------------------|
| 1      | 6246   | 88.7        | 1416   | 4 09B8F9 | 09b8f9 homo sapien |
| 2      | 6190   | 87.9        | 1449   | 4 09B8F7 | 09b8f7 homo sapien |
| 3      | 6056   | 86.0        | 1388   | 4 09G121 | 09g121 canis famli |
| 4      | 5027   | 71.4        | 1151   | 4 09HCL1 | 09hcl1 homo sapien |
| 5      | 1322.5 | 11.9        | 983    | 4 09NWC6 | 09nwc6 homo sapien |
| 6      | 1205   | 11.7        | 980    | 4 09P2L2 | 09p2l2 homo sapien |
| 7      | 985.5  | 14.1        | 1227   | 4 09P6K7 | 09p6k7 mus sapien  |
| 8      | 985.5  | 14.1        | 1227   | 4 09P6K7 | 09p6k7 mus sapien  |
| 9      | 985.5  | 14.1        | 1227   | 4 09P6K7 | 09p6k7 mus sapien  |
| 10     | 655.5  | 9.3         | 1710   | 4 09H101 | 09h101 mus sapien  |
| 11     | 654    | 9.3         | 1709   | 4 09P9S8 | 09p9s8 homo sapien |
| 12     | 640.5  | 9.1         | 1410   | 4 014221 | 014221 homo sapien |
| 13     | 640.5  | 9.1         | 2779   | 5 09M4N7 | 09m4n7 drosophila  |
| 14     | 628    | 8.9         | 1411   | 4 015075 | 015075 homo sapien |
| 15     | 621.5  | 8.8         | 1531   | 4 091973 | 091973 cornuix co  |
| 16     | 618    | 8.8         | 1526   | 4 015085 | 015085 homo sapien |
| 17     | 618    | 8.7         | 1930   | 6 092960 | 092960 saccharomyc |
| 18     | 613    | 8.7         | 1935   | 6 09GCR1 | 09gcr1 saccharomyc |
| 19     | 610.5  | 8.7         | 2139   | 5 007569 | 007569 entamoeba h |

| RESULT | 1                                                            | ALIGNMENTS |
|--------|--------------------------------------------------------------|------------|
| 09B2F9 | PRELIMINARY: PRT; 1416 AA.                                   |            |
| ID     | 09B2F9                                                       |            |
| AC     | 09B2F9                                                       |            |
| DT     | 01-JUN-2001 (PREMREL, 17, Created)                           |            |
| DR     | 01-JUN-2001 (PREMREL, 17, Last sequence update)              |            |
| DE     | 01-JUN-2001 (PREMREL, 17, Last annotation update)            |            |
| DE     | UPBAL, AUTOANNOTIGEN                                         |            |
| OS     | Homo sapiens (Human).                                        |            |
| OC     | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;      |            |
| OC     | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.    |            |
| OX     | NCBI_TaxID=9606;                                             |            |
| RN     | [1]                                                          |            |
| RF     | SEQUENCE FROM N.A.                                           |            |
| RA     | PubMed=1162650;                                              |            |
| RA     | Nakamura K., Senji S., Nakamura T., Murata Y., Ishihara K.,  |            |
| RT     | "Identification of a novel autoantigen UbcA in patients with |            |
| RT     | panuveitis."                                                 |            |
| RL     | Biochem. Biophys. Res. Commun. 280:1169-1176(2001).          |            |
| DR     | EMBL: AF322916; AAC49577.1;                                  |            |
| SO     | SEQUENCE 1416 AA; 162579 MW; DAE2F5B5AA76BD CRC64;           |            |

| Query Match           | 88.7%                                                         | Score 6246                                     | DB 4                      | Length 1416 |
|-----------------------|---------------------------------------------------------------|------------------------------------------------|---------------------------|-------------|
| Best local similarity | 88.7% <td>Score 6246 <td>DB 4 <td>Length 1416</td> </td></td> | Score 6246 <td>DB 4 <td>Length 1416</td> </td> | DB 4 <td>Length 1416</td> | Length 1416 |
| Matches 1333          | Conservative 84                                               | Mismatches 74                                  | Indels 4                  | Gaps 2      |
| 9                     | AAKRNQADWKNKDYDRLLAKRAKRDYKVSIIIAKKGNGKLVDSRGRVHVAASKN        | 68                                             |                           |             |
| 22                    | AAASVRAADWKNKDYDRLLAKRAKRDYKVSIIIAKKGNGKLVDSRGRVHVAASKN       | 81                                             |                           |             |
| 69                    | ECVALLILHGVDTTSDTACRRAALHLAAVGHACLDLQITNCPTEHYDQGRTRALDA      | 128                                            |                           |             |
| 82                    | ECVALLILHGVDTTSDTACRRAALHLAAVGHACLDLQITNCPTEHYDQGRTRALDA      | 141                                            |                           |             |
| 129                   | AAACCPSTSIOLCDHGASVNAKVDGPRFLVLAQWKCPPTICQLLIDRQADINSRDKRN    | 188                                            |                           |             |
| 142                   | AAACCPSTSIOLCDHGASVNAKVDGPRFLVLAQWKCPPTICQLLIDRQADINSRDKRN    | 201                                            |                           |             |
| 189                   | TAAACGCGKQAYEVILKNCADVTLTDLALGHSSTYATIDRDLITLTALKASNSNK       | 248                                            |                           |             |

```

Db 202 TALMGCEGCDADVILKMGADISLLOGLDHSSYARIGONIDITLTLKTSSENTNK 261
Oy 249 GRELLKKGPSLQOORNLDSOMLDEYVNRKSNOREHONITDOLIEINENDDLEKRLKTOOEORILL 308
Db 262 GRELLKKGPSLQOORNLTHMOEYVNRKSNOREHONITDOLIEINENDDLEKRLKTOOEORILL 321
Oy 309 DKNVGLQOLQJLNEEYVADLESEKRLKSLLAKEQKHESSLTTEBALKSRKFFESDHL 368
Db 322 DKNVGLQOLQJLNEEYVADLESEKRLKSLLAKEQKHESSLTTEBALKSRKFFESDHL 381
Oy 369 GSGSHF--RKEDMLLKOGOMYNTDSCSTGCPYHMOSSMLPRLTELALPNQASYSRNEI 426
Db 382 GSGSHFSNRKEDMLLKOGOMYVADSCSTGCPYHMOSSMLPRLTELALPNQASYSRNEI 441
Oy 427 LKKELEAMRTFCDASAKODRLKONELAHKVAACALALECERKVEDSDIOIKOLEDALKD 486
Db 442 LKKELEAMRTFCDASAKODRLKONELAHKVAACALALECERKVEDSDIOIKOLEDALKD 501
Oy 487 VOKRRYTESBCKYKQOMOTHFALKEILTSDAAGNHRIMLEKIPKOLKDMKYYKEASAEVG 546
Db 502 VOKRRYTESBCKYKQOMOTHFALKEILTSDAAGNHRIMLEKIPKOLKDMKYYKEASAEVG 561
Oy 547 KLANOIKONEMLYVEERDECKLMENNRKLORELKSCLELREKRRKTLTENGOLKDLISA 606
Db 562 KLANOIKONEMLYVEERDECKLMENNRKLORELKSCLELREKRRKTLTENGOLKDLISA 621
Oy 607 KLAISTPAKPEFNKSSLSNETMEYAKKLLDYEPESRSLNORLKRLEPMYAKKIAQH 666
Db 622 KLAISTPAKPEFNKSSLSNETMEYAKKLLDYEPESRSLNORLKRLEPMYAKKIAQH 681
Oy 667 VKREHEDKSLRLQKSGELCKRITELTSGNKTLOKRIEKCVCJDNKLLTTOYVNMLTTEKK 726
Db 682 VKREHEDKSLRLQKSGELCKRITELTSGNKTLOKRIEKCVCJDNKLLTTOYVNMLTTEKK 741
Oy 727 N--YOLKYSBCKKSDYVTDYAKKLSIDYHRTKTEKLLPCKGLKEMASLKNYSMET 784
Db 742 NHRVLYKSLDKKSDIDAIIDDNLKRLDVTQKTEKLEHKLLENDSTLKNVSHMET 801
Oy 785 VFPREPHEKEMALKSNTTELKOLSKSLKCCBDEYITSLMSLNDKATKTSHQYAP 844
Db 802 VFPREPHEKEMALKSNTVELKOLSKSLKCCBDEYITSLMSLNDKATKTSHQYAP 861
Oy 845 VKTBEIKTALSTLQKTNRELVDYKKCCDINOEFYKIKDENEILKRNLENTONVYAE 904
Db 862 VKTBEIKTALSTLQKTNRELVDYKKCCDINOEFYKIKDENEILKRNLENTONVYAE 921
Oy 905 YLSLRHEKMGSLRKSMKAYOONSLELAKKYSKQSVYUJLHEJLAOKRELOTOIOEI 964
Db 922 YLSLRHEKMGSLRKSMKAYOONSLELAKKYSKQSVYUJLHEJLAOKRELOTOIOEI 981
Oy 965 KLYKAPITSLSECEKPKATERELKLBOLSOOTQYVNTSEBAPKKOENKLEKLETLQ 1024
Db 982 KLYKAPITSLSECEKPKATERELKLBOLSOOTQYVNTSEBAPKKOENKLEKLETLQ 1041
Oy 1025 KDLKOKVNHLENSYFETPRALSRKTEPLNORLQDLQYKFAKKERKLEKLENAKOTSEIL 1084
Db 1042 KDLKOKVNHLENSYFETPRALSRKTEPLNORLQDLQYKFAKKERKLEKLENAKOTSEIL 1101
Oy 1085 AAOITLLOKHVPLDQVESLKSLSGTEFLKLEELTKQKCEKEDQOYVTOLOKMLENOMN 1144
Db 1102 AAOITLLOKHVPLDQVESLKSLSGTEFLKLEELTKQKCEKEDQOYVTOLOKMLENOMN 1161
Oy 1145 SSYVLAHEDVKEAFKEVGTITNAISLRKREESONKTEBESKQASFTONTKQALKLETR 1204
Db 1162 SSYVLAHEDVKEAFKEVGTITNAISLRKREESONKTEBESKQASFTONTKQALKLETR 1221
Oy 1205 EYVDSLKTKTKSDLETOISDLNENKLANLRKYEEVCEVYVLAHKKKSLSKOKKELHHS 1264
Db 1222 EYVDSLKTKTKSDLETOISDLNENKLANLRKYEEVCEVYVLAHKKKSLSKOKKELHHS 1281
Oy 1265 IEOETIKDOQERDONSLTITTELORHIOSSAKOITAKDNKTITELLNDYERLKLQALNGLSOL 1324
Db 1282 IEOETIKDOQERDONSLTITTELORHIOSSAKOITAKDNKTITELLNDYERLKLQALNGLSOL 1341

```

```

Oy 1325 TYGSGSPSKRSKOSLSDLOOQVARSLOOGLADADQROHONVETIARTHLLSAMQGHMDEVO 1384
Db 1342 TYGSGSPSKRSKOSLSDLOOQVARSLEOGLADADQROHONVETIARTHLLSAMQGHMDEVO 1401
Oy 1385 AALLQITQNRQGLVC 1399
Db 1402 EALLQITQNRQGLVC 1416

RESULT 2.
Oy 09BGR7 PRELIMINARY; PRT: 1449 AA.
AC 09BGR7;
DT 01-JUN-2001 (TTEMBJREL 17, Created)
DT 01-JUN-2001 (TTEMBJREL 17, Last sequence update)
DE UYVAL APOANTIERK
OS Boia taurus (Bovinae).
OC Buarqueiros; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae.
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RP Yarnada K., Scott S., Nakamura T., Kurata Y., Ishihara M.,
RA Nakamura S., Ohno S., Negi A., Nishimura Y.;
RT Identification of a novel autoantigen UCA4 in patients with
RL pancytopenia.
DR Biochem. Biophys. Res. Commun. 280:1169-1176(2001).
EMBL: AF322915; AAC49576.1; -.
SO SEQUENCE 1449 AA; 165881 MW; 7A73CA4AECBA814C CRC64;

Query Match 87.9%; Score 6190; DB 6; Length 1449;
Basic Local Similarity 88.3%; Pred. No. 6,7e-214;
Matches 1222; Conservative 84; Mismatches 74; Indels 4; Gaps 2;

```



|                                   |                                                     |                                                              |      |
|-----------------------------------|-----------------------------------------------------|--------------------------------------------------------------|------|
| Db                                | 502                                                 | VORRYSEGGYKONOTHFALAKKHLITSEASNGHRLTBELOKQOLKLYKRYEGASAEVQ   | 561  |
| Oy                                | 547                                                 | KLRNIOIKONEMLYVEFRKREKRLMEKRRKLORELSNCELRERKRLKTYEMKCOLKDISA | 606  |
| Ox                                | 552                                                 | KLRNIOIKONEMLYVEFRKREKRLMEKRRKLORELSNCELRERKRLKTYEMKCOLKDISA | 621  |
| Db                                | 607                                                 | KLAUSIPAEKFNKSKLSLSNELMEKAKKLIDVEREYSLNETRPLKRELENLAKLAQ     | 666  |
| Db                                | 622                                                 | KLAUSIPAEKFNKSKLSLSNELMEKAKKLIDVEREYSLNETRPLKRELENLAKLAQ     | 661  |
| Oy                                | 667                                                 | VKREHEHOKVSKLBOKSGELKRLITELTSKNTQLOREIEXYCDKRLTDOVNNKLTPEK   | 726  |
| Db                                | 682                                                 | VKREHEHOKVSKLBOKSGELKRLITELTSKNTQLOREIEXYCDKRLTDOVNNKLTPEK   | 741  |
| Oy                                | 727                                                 | N-VELKVESEKRSADVVDLNRKLSVDHAKTKRKLEKLEMLKMSLSKVSLET          | 784  |
| Db                                | 742                                                 | NHVLKLVSEDKKSDALIDDLNRKLLDVOTKTKRKLEKLEMLKMSLSKVSLET         | 801  |
| Oy                                | 785                                                 | VFLPPEHEKEMALNTITELKNTKOLSELKNGKOEKXTYSLSKNSNDIKYMSNOYRP     | 844  |
| Db                                | 802                                                 | VFLPPEHEKEMALNTITELKNTKOLSELKNGKOEKXTYSLSKNSNDIKYMSNOYRP     | 861  |
| Oy                                | 845                                                 | VKTHREKTKALSSTLDKTNRELVDKPKKCEDINOEFVKIKDEMLKRLNLTONTONVKE   | 904  |
| Db                                | 862                                                 | VKTHREKTKALSSTLDKTNRELVDKPKKCEDINOEFVKIKDEMLKRLNLTONTONVKE   | 921  |
| Oy                                | 905                                                 | YSLAEHEKMSGLRKSKMYODNSALITAKYKSGEETVLAEBLAKOKELOTOCT         | 964  |
| Db                                | 922                                                 | YSLAEHEKMSGLRKSKMYODNSALITAKYKSGEETVLAEBLAKOKELOTOCT         | 981  |
| Oy                                | 965                                                 | KLKVAPRILSLSECRKFKATKELKEOLSOOTQKRYTSEBAKCKOENDKILKEITLO     | 1024 |
| Db                                | 982                                                 | KLKVAPRILSLSECRKFKATKELKEOLSOOTQKRYTSEBAKCKOENDKILKEITLO     | 1041 |
| Oy                                | 1025                                                | KOLKORVNIENYSEPRALSKREBELANQLKLOLTKTTAKKEREKLYVENAKQISL      | 1084 |
| Db                                | 1042                                                | KOLKORVNIENYSEPRALSKREBELANQLKLOLTKTTAKKEREKLYVENAKQISL      | 1101 |
| Oy                                | 1085                                                | MAOTLQKOHVPLEOYESLKLSSGTELETKLEIKKOCYKREOQVYTOLOKLEMKRN      | 1144 |
| Db                                | 1102                                                | MAOTLQKOHVPLEOYESLKLSSGTELETKLEIKKOCYKREOQVYTOLOKLEMKRN      | 1161 |
| Oy                                | 1145                                                | SSVPLAEHLQKFAPEKXEVQIKASLAKREBESQNTVEYSKLOSETONTOKLKLLETR    | 1204 |
| Db                                | 1162                                                | SSVPLAEHLQKFAPEKXEVQIKASLAKREBESQNTVEYSKLOSETONTOKLKLLETR    | 1221 |
| Oy                                | 1205                                                | EYVDSLAKYKATSDLETOISLNEKLANIKRYEYCEVLAHKKELSKMEKELLES        | 1264 |
| Db                                | 1222                                                | EYVDSLAKYKATSDLETOISLNEKLANIKRYEYCEVLAHKKELSKMEKELLES        | 1281 |
| Oy                                | 1265                                                | IRDEIKQOECRQNSLTTTLELORIOESNAKOKTEAKKTETELLANDYERLAKALNSLQ   | 1324 |
| Db                                | 1282                                                | IRDEIKQOECRQNSLTTTLELORIOESNAKOKTEAKKTETELLANDYERLAKALNSLQ   | 1341 |
| Oy                                | 1325                                                | TYGSSPSKROSLDLSLOOVASLOOQLADNDHOEYIAFRHLLSLMAAGHNOEDVQ       | 1384 |
| Db                                | 1342                                                | TYGSSPSKROSLDLSLOOVASLOOQLADNDHOEYIAFRHLLSLMAAGHNOEDVQ       | 1401 |
| Oy                                | 1385                                                | ALL 1388                                                     |      |
| Db                                | 1402                                                | ALL 1405                                                     |      |
| RESULT 3                          |                                                     |                                                              |      |
| 09G121 PRELIMINARY: PRT: 1388 AA. |                                                     |                                                              |      |
| Id                                | 09G121                                              | PRELIMINARY: PRT: 1388 AA.                                   |      |
| Ac                                | 01-MAR-2001 (Tributrel, 16, Created)                |                                                              |      |
| Df                                | 01-MAR-2001 (Tributrel, 16, Last sequence update)   |                                                              |      |
| Dp                                | 01-MAR-2001 (Tributrel, 17, Last annotation update) |                                                              |      |
| GN                                | C3VS. PROTEIN.                                      |                                                              |      |
| OS                                | Canis familiaris (dog).                             |                                                              |      |

|    |                                                                     |
|----|---------------------------------------------------------------------|
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |
| OC | Mammalia; Eutheria; Carnivora; Placentalia; Canidae; Canis.         |
| Ox | Ncbi_Taxid=9615;                                                    |
| RP | SEQUENCE FROM N. A.                                                 |
| RC | SEQUENCE FROM N. A.                                                 |
| RA | WILKIN F., Savonnet R., Radulescu A., Petermans J., Dumont J.E.,    |
| RT | Identification and Characterization of Novel Genes Modulated in the |
| RL | Thyroid of Dogs Treated with Methimazole and Propylthiouracil."     |
| RM | J. Biol. Chem. 271:28451-28457(1996).                               |
| RP | SEQUENCE FROM N. A.                                                 |
| RA | Savonnet V.;                                                        |
| RL | Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.             |
| RM | EMBL: X59145; Accession: 1.                                         |
| DR | Protein: P00023; ank: 6.                                            |
| DR | SMART: SM00248; ANK: 6.                                             |
| DR | PROSITE: PS50088; ANK_REPEAT: 5.                                    |
| DR | PROSITE: PS50297; ANK_REGION: 1.                                    |
| SO | SEQUENCE 1388 AA; 159459 MW; 9A50BD2754A9D6 CRC64;                  |

|               |                   |                  |            |              |
|---------------|-------------------|------------------|------------|--------------|
| Query Match   | 86.08;            | Score 6056;      | DB 5;      | Length 1388; |
| Identity      | 74.74;            | Field No. 9e-20; |            |              |
| Matches 1186; | Conservative 108; | Mismatches 977;  | Indels 30; | Gaps 2;      |

|    |     |                                                           |     |
|----|-----|-----------------------------------------------------------|-----|
| Oy | 1   | MMSCWFCAPKRNROADNKKYDRLMRARADVDYEVSSILAKKCVNFKLVGSGSAFH   | 60  |
| Db | 16  | MMSCWFCAPKRNROADNKKYDRLMRARADVDYEVSSILAKKCVNFKLVGSGSAFH   | 75  |
| Oy | 61  | VVASGMLCELMALIHGVDTTSDTNGRNALMLAAKVGKALCTOKLLOTCPTPEVLDQ  | 120 |
| Db | 76  | VVASGMLCELMALIHGVDTTSDTNGRNALMLAAKVGKALCTOKLLOTCPTPEVLDQ  | 135 |
| Oy | 121 | GRTALHDLAAKAPSSITOLCDHIGASVNAVDGRTPLVATOKCRPTICOLLIDRGAD  | 180 |
| Db | 136 | GRTALHDLAAKAPSSITOLCDHIGASVNAVDGRTPLVATOKCRPTICOLLIDRGAD  | 195 |
| Oy | 181 | NSSDKRNTRTALMLCCGYCGKAVAYELKNGADVTLLDALGDSSTYARIGNDLITLTK | 240 |
| Db | 196 | NSSDKRNTRTALMLCCGYCGKAVAYELKNGADVTLLDALGDSSTYARIGNDLITLTK | 255 |
| Oy | 241 | TASEKSKNGKRELMKPSLOORNLKSLMDPVNTSNGRPHONITOLLEKEDLERKRT   | 300 |
| Db | 256 | TASEKSKNGKRELMKPSLOORNLKSLMDPVNTSNGRPHONITOLLEKEDLERKRT   | 315 |
| Oy | 301 | QOEBRILILDVNGLOLNEVYVADLSESEKELSLLAKEKREHESLRTITALKSRF    | 360 |
| Db | 316 | QOEBRILILDVNGLOLNEVYVADLSESEKELSLLAKEKREHESLRTITALKSRF    | 375 |
| Oy | 361 | KRYESDHLGSGHFRKEDMLKQGGOMVTDISQCTSGPMYHMOSSMLPTELALPKQAS  | 420 |
| Db | 376 | KRYESDHLGSGHFRKEDMLKQGGOMVTDISQCTSGPMYHMOSSMLPTELALPKQAS  | 435 |
| Oy | 421 | YSENLILKLELMARFTCDASAKODRLKONELAKVAEBKALALCEDEVEEDSDQIKOL | 480 |
| Db | 436 | YSENLILKLELMARFTCDASAKODRLKONELAKVAEBKALALCEDEVEEDSDQIKOL | 495 |
| Oy | 481 | EDLADKQKRWYSEGGYKQKOTHFALAKKHLITSEASNGHRLTBELOKQOLKLYKRY  | 540 |
| Db | 496 | EDLADKQKRWYSEGGYKQKOTHFALAKKHLITSEASNGHRLTBELOKQOLKLYKRY  | 555 |
| Oy | 541 | ASAEVQKLRNOKNEMLYVEFRKREKRLMEKRRKLORELSNCELRERKRLKTYEMK   | 600 |
| Db | 528 | ASAEVQKLRNOKNEMLYVEFRKREKRLMEKRRKLORELSNCELRERKRLKTYEMK   | 587 |
| Oy | 601 | KLAUSIPAEKFNKSKLSLSNELMEKAKKLIDVEREYSLNETRPLKRELENLAKLAQ  | 660 |
| Db | 598 | KLAUSIPAEKFNKSKLSLSNELMEKAKKLIDVEREYSLNETRPLKRELENLAKLAQ  | 647 |

```

661 AKLAQHVAPEEHDLKSRLEQSGCELRITETSKNOCTLOKLEIEKVLONKLLTQOVNN 720
648 AKLAQHVAPEEHDLKSRLEQSGCELRITETSKNOCTLOKLEIEKVLONKLLTQOVNN 707
721 LTTENK--VPLAKSEPKKSHOYIVYDOLKALSDYHKTETKLEKLEKLEKLEKLEK 778
708 LTSEKSHVYPLQVSEBKSHOYVTEELKOLQDLYOTQOCADQLOMEKLLLENDLSIKN 767
779 VSRLEFVPIPEHNEKEMALKSITELKQSLSEKNGCEQOKTYSYLSMEQNDKKTN 838
766 VSRLEFVPIPEHNEKEMALKSITELKQSLSEKNGCEQOKTYSYLSMEQNDKKTN 827
839 SHOYVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 898
828 SHOYVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 887
899 NOVAKVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 958
888 SOLKABYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 947
959 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1018
948 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1007
1019 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1078
1008 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1067
1079 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1138
1068 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1127
1139 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1187
1128 TIOECIKLAPVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1187
1199 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1258
1188 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1247
1259 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1318
1248 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1307
1319 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1378
1308 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1367
1379 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1399
1368 KLEFREVYVYVTEHETKALSTLDKTNELVYKCCEDINOEFVAKDSNEIKLEKLENT 1388

```

```

RT code for large proteins in vitro.
RT DNA Res. 7:273-281(2000).
DR EMBL; AB046781; BAB13387.1;
DR InterPro; IPR000727; T_SNRK.
FT NON-TER
SQ SEQUENCE 1151 AA; 13404 MW; AFA8591C3E20A508 CRC64;

Query Match 71.4%; Score 5027; DB 4; Length 1151;
Match local similarity 87.4%; Pct. 1.9e-17;
Matches 1001; Conservative 77; Mismatches 63; Indels 4; Gaps 2;

259 LQQRNLSQMDIEVYKSNQREHODIDLEENEDKLEKRLKIQDEORTLLDQVNGVLDL 318
7 LKRLNLTDMODEVYKSNQREHODIDLEENEDKLEKRLKIQDEORTLLDQVNGVLDL 66
319 NQYV 376
67 NQYV 126
377 EDMLKQOQVYV 436
127 EDMLKQOQVYV 186
437 FQDSAKQRLKQLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 496
187 FQDSAKQRLKQLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 246
497 KVMQKQTHFLAKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 556
247 KVMQKQTHFLAKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 306
557 KVMQKQTHFLAKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 616
307 KVMQKQTHFLAKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 365
617 FQDSAKQRLKQLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 676
367 FQDSAKQRLKQLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 426
677 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 734
427 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 486
735 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 794
487 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 546
795 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 854
547 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 606
855 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 914
607 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 666
915 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 974
667 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 726
975 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 1034
727 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 786
1035 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 1094
787 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 846
1095 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 1154
847 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 906
1155 SRLEKQSGELKRLTETSKNOCTLOKLEIEKVLONKLLTQOVNNLTTEPMKN--VPLK 1214

```

|    |      |                                                                |       |
|----|------|----------------------------------------------------------------|-------|
| Db | 907  | IKFAEFKEVGIITASJPRKEEESNMKEEVSKLQSVQNTQALMKTREVDLSEKTKA        | 966   |
| Qy | 1215 | TKSDLEFQISQLENEKLANLRKRYEVECEVYLAKKKELLSAKDEKLLHSIEDEIRIDOE    | 1274  |
| Db | 967  | TKSDLEFQISQLENEKLANLRKRYEVECEVYLAKKKELLSAKDEKLLHSIEDEIRIDOE    | 10264 |
| Qy | 1275 | KCDLSLTITTELQRIQESAKQIEAKDNKTITELLNDYVERIKQAQLNGLSQTVGS6SGP8KR | 13334 |
| Db | 1027 | KCDLSLTITTELQRIQESAKQIEAKDNKTITELLNDYVERIKQAQLNGLSQTVGS6SGP8KR | 10866 |
| Qy | 1335 | QSGSLISLQOQYSLQDQADPDQHQHGVIAITPHTLLSAGQIHMDYQALALQITOMR       | 13964 |
| Db | 1087 | QSGSLIDPDQHQVSTLEQGLADPDQHQHGVIAITPHTLLSAGQIHMDYQALALQITOMR    | 11466 |
| Qy | 1395 | QGLVNC 1399                                                    |       |
| Db | 1147 | QGLVNC 1151                                                    |       |

|          |                                                                    |                                         |              |  |  |
|----------|--------------------------------------------------------------------|-----------------------------------------|--------------|--|--|
| RESULTS  | 5                                                                  |                                         |              |  |  |
| 09NMNC6  |                                                                    |                                         |              |  |  |
| ID       | 09NMNC6                                                            | PRELIMINARY:                            | PRT: 535 AA. |  |  |
| AC       | 09NMNC6                                                            |                                         |              |  |  |
| DT       | 01-OCT-2000                                                        | (TREMBLrel. 15, Created)                |              |  |  |
| DF       | 01-OCT-2000                                                        | (TREMBLrel. 15, Last sequence update)   |              |  |  |
| DE       | 01-OCT-2000                                                        | (TREMBLrel. 15, Last annotation update) |              |  |  |
| OS       | HYPOCRETINIC PROTEIN (PROMOTIN) .                                  |                                         |              |  |  |
| DS       | Homo sapiens (human) .                                             |                                         |              |  |  |
| OC       | Eukaryotes; Metazoa; Chordata; Vertebrata; Euteleostomi;           |                                         |              |  |  |
| CC       | Mammalia; Eutheria; Placental; Carnivora; Homiidae; Homo.          |                                         |              |  |  |
| OX       | NCBI_TaxID:9606;                                                   |                                         |              |  |  |
| RS       | SEQUENCE FROM N.A.                                                 |                                         |              |  |  |
| RC       | TIGS1288EMHMO;                                                     |                                         |              |  |  |
| RA       | Issigal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., |                                         |              |  |  |
| RA       | Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,          |                                         |              |  |  |
| RA       | Megatsuma M., Hosoi T., Kaku Y., Kodaka H., Sugawara H.,           |                                         |              |  |  |
| RA       | Takahashi S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takijunchi |                                         |              |  |  |
| RA       | McLachlan S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K. |                                         |              |  |  |
| RA       | Nimmoja K., Iwanaga T., Makamura T., Nagahari K., Masuhio Y.,      |                                         |              |  |  |
| RT       | "NEO human cDNA sequencing project."                               |                                         |              |  |  |
| RL       | Submitted (FEB-2000) to the EMBL/Genbank/DBD databases.            |                                         |              |  |  |
| EMBL:    | AK000990; BA01571.1; "                                             |                                         |              |  |  |
| NC       | 535                                                                | 535                                     |              |  |  |
| NON_TER  |                                                                    |                                         |              |  |  |
| SEQUENCE | 535 AA; 62604 MM; C5C0CF36852791101 CRC64;                         |                                         |              |  |  |

| Query Match           | 31.5%           | Score 2201                                                    | De 4     | Length 535 |
|-----------------------|-----------------|---------------------------------------------------------------|----------|------------|
| Best Local Similarity | 11.9%           | Pred. No. 716-72                                              |          |            |
| Matches 438           | Conservative 48 | Mismatches 47                                                 | Indels 2 | Gaps 1     |
| Qy                    | 638             | VREYERSEINTEPRKLEINIKAKIOMVYREHDOIKREYDQKSEGRITRELTYSKN       | 637      |            |
| Dy                    | 1               | MREHREISSEISITKOLKRELENNKAKLAOHVPRHEHVOYSKLEQSGDEKLTITELTN    | 60       |            |
| Qy                    | 698             | OTLOKEITKLVCLNNKLLTPOGVNNTTDEKN--VPLKSEEMKSHDVIYDULNNKLSLQV   | 755      |            |
| Dy                    | 61              | OTLOKEITKLVCLNNKLLTPOGVNNTTDEKNVPLKSEEMKSHDVIYDULNNKLSLQV     | 120      |            |
| Qy                    | 756             | KHYEKKELMEKLLENASVLSKSNYSRLIEVPIPRHEHREEMAKLSNTEKLSKOLSELKN   | 815      |            |
| Dy                    | 121             | OKYETREKLEMEKLTLENDSVDSRLIEVPRHEHREKSIITAKSNISYEKLSKOLSELK    | 180      |            |
| Qy                    | 816             | KQKDEDKETSLASNDLKTAKSHQVIVYATETIEETALSSTLDKTNELVDYKACED       | 875      |            |
| Dy                    | 181             | KQKDEDKETALSSTNRLKAKSHQVIVYATETIEETALSSTLDKTNELVDYKACED       | 240      |            |
| Qy                    | 876             | INOEFTYKIDENSIILKRLLENTOQVAVATYSILREHEKKSOLDKSRMYKVDONSACILTA | 935      |            |
| Dy                    | 241             | INOEFTYKIDENSIILKRLLENTOQVAVATYSILREHEKKSOLDSKMYKVDONSACILTA  | 300      |            |

|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| 0Y | 936  | KASODEIYVILBEI LAOKRELLTOIGERIKKVAPII SLEBCEBRRPAAFEKELKOL500 | 936  |
| 0Y | 937  | KASODEIYVILBEI LAOKRELLTOIGERIKKVAPII SLEBCEBRRPAAFEKELKOL500 | 937  |
| 0Y | 301  | KOKODEIYVILBEI LAOKRELLTOIGERIKKVAPII SLEBCEBRRPAAFEKELKOL500 | 360  |
| 0Y | 996  | TOXNENNEEAAKOKMOKDOKKEI LKTELLOKXKXNENENSYEPALISPTREKOL500    | 105  |
| 0Y | 997  | TOXNENNEEAAKOKMOKDOKKEI LKTELLOKXKXNENENSYEPALISPTREKOL500    | 105  |
| 0Y | 361  | TOXSYSEEVKKNKNOBKKELEFTELLOKXKXNENENSYEPALISPTREKOL500        | 420  |
| 0Y | 1056 | KOLLOJYEPAAKEEKEIKVENAAKOTSEILAQTLOKOHPELOVSELSKSLGCTIEFK     | 1111 |
| 0Y | 421  | KOLSOJYEPAAKEEKEIKVENAAKOTSEILAQTLOKOHPELOVSELSKSLGCTIEFK     | 480  |
| 0Y | 1116 | BEIKTOKRCEKROGTIVTLOKLEKNOKNSVPALEHJOVKEAPEKGVGIKRSI          | 1170 |
| 0Y | 481  | BEIKSOMOKCEKROGTIVTLOKLEKNOKNSVPALEHJOVKEAPEKGVGIKRSI         | 535  |

| Q  | RESULT                                  | 6                                                                                    |                           |
|----|-----------------------------------------|--------------------------------------------------------------------------------------|---------------------------|
| 1  | 099212                                  |                                                                                      | PRELIMINARY: PRT: 989 AA. |
| 2  | 099212                                  |                                                                                      |                           |
| 3  | 01-OCT-2000                             | (TREMURBL. 15, Created)                                                              |                           |
| 4  | 01-OCT-2000                             | (TREMURBL. 15, Last sequence update)                                                 |                           |
| 5  | 01-JUN-2001                             | (TREMURBL. 17, Last annotation update)                                               |                           |
| 6  | KIAA1334                                | PROTEIN (FRAGMENT).                                                                  |                           |
| 7  | KIAA1334                                |                                                                                      |                           |
| 8  | Non sapiens (human)                     | (Chordata: Vertebrata: Euteleostomi:                                                 |                           |
| 9  | Homio                                   | Actinopterygii: Cyprinodontiformes: Poeciliidae:                                     |                           |
| 10 | Mammalia: Eutheria                      | : Primates: Catarrhini: Hominoidea: Homo:                                            |                           |
| 11 | NCBI_Taxid=9606;                        |                                                                                      |                           |
| 12 | [1]                                     |                                                                                      |                           |
| 13 | SEQUENCE FROM N.A.                      |                                                                                      |                           |
| 14 | TSUSSU-BEAM:                            |                                                                                      |                           |
| 15 | NCBI_Taxid=7018126;                     | PubMed=10718189;                                                                     |                           |
| 16 | NCBI_Taxid=7018126;                     | Hirosewa M., Ohara O.,                                                               |                           |
| 17 | NCBI_Taxid=7018126;                     | Matsuda T. (1999) Prediction of the coding sequences of unidentified human genes XVI |                           |
| 18 | RT                                      | The complete sequences of 150 new cDNA clones from brain which code                  |                           |
| 19 | RT                                      | for large proteins in vitro.                                                         |                           |
| 20 | RNA Res. 7:65-73(2000).                 |                                                                                      |                           |
| 21 | EMBL: AB037755; BA92572.1;              |                                                                                      |                           |
| 22 | RefSeq: NM_002042.1; ANK.               |                                                                                      |                           |
| 23 | RefSeq: NM_002042.1; ANK.               |                                                                                      |                           |
| 24 | SMART: SM00248; ANK: 5.                 |                                                                                      |                           |
| 25 | PROSITE: PS50088; ANK_REPEAT: 5.        |                                                                                      |                           |
| 26 | PROSITE: PS50297; ANK_REPEAT_REGION: 1. |                                                                                      |                           |
| 27 | NON_TER                                 |                                                                                      |                           |
| 28 | SEQUENCE                                | 989 AA: 11020 MW: EA84FB9926E23JFS CRC64:                                            |                           |

|    | Query Match                                                        | 17.8%: Score 1252.5; Db 4; Length 989; |
|----|--------------------------------------------------------------------|----------------------------------------|
|    | Best Local Similarity                                              | 26.1%: Pred. No. 1.1e-37;              |
|    | Matches 37; Conserved/ity 211; Mismatches 332; Indels 507; Gaps    | 33.3%                                  |
| Oy | 17 PMNXYDPLDAAEALRENGDUEKSSALKGVNCPKCDVEGRSAFVYASQCEGLDCLAI 76     |                                        |
| Db | 24 ENKNDNDRLDVAENGDAEKVSLIKLAKASAKRHDSEKATFHLMAAGVCELRWIT 83       |                                        |
| Oy | 77 HGVDTTSPDTAGNNALHLAAKKGALCLQKLDLYNCPPEHVLDDGRTALHDAAMACSS 136   |                                        |
| Db | 84 HGVVDVDDDTGTGSHALHLAAKSSHREKTRLLDGRKPEVSSSKETALHAAQCCIA 143     |                                        |
| Oy | 137 TQLDLCGASVNAKVGSGRTPLVATQMCQRPETCOLLDLDGADINSNDQRPALMLCGE 196  |                                        |
| Db | 144 VOICEKSPINKLDCDNPILLALVANGHSEICHELDDIGADVNSRNSGRATMLACE 203    |                                        |
| Oy | 197 YCCDAVEVLLKKGADVTLLDALHDSSTVYARIGDIDLITLL 239                  |                                        |
| Db | 204 TQSSNVEALLKKGADLVKQDADLQNLNLRKSLSEMGDQSLSDQDADKPTPTFR 263      |                                        |
| Oy | 240 TQSSNVEALLKKGADLVKQDAGSTLQNLNLRKSLSEMGDQSLSDQDADKPTPTFR 293    |                                        |
| Db | 264 OHQDVAKTSSERGGTPKKRKAAPRPIISPTQLDSVSPASTTSTLQSK--ESYFAVEPP 321 |                                        |

```

Oy 294 KERUKTOOPOBILLDKVNGLQLOLNEBVMVADL-----PSEKIKSLAAKERO--- 345
Db 322 KAEIISIREKROILSDSTTG-----ADSLDISBDAODLISLDAKVASLTL 370
Oy 346 HESILPTIEALSKRFKFEPSDHLGSGHFHFKEDM---LLKOGOMINTDSQCTSGNPFVIM 402
Db 371 HNEELQDLQAKSP-KAEADLSFDSYSTRVDLGPRLQKCKGCTSPDSK-SSPSVLIHS 428
Oy 403 OSHSHAPLELALPNOASTSENELIKLELDMARFCDKADRIKONELAIKVAECKAL 462
Db 429 LGMS----- 432
Oy 463 ALCEGVNEDSDQIKQLEDAIKDVQKHWYSESEKVCYKQNHTEHLALKEHLTSDAQTGNHR 522
Db 433-----TTDNVVRIDQLOETLQDLQKLTSSAEKROLO----- 465
Oy 523 LMBELKQDLQAKKRYEGASGVGLKMNQKONEMLEVEFRDEQKLMENKRIQKLELKM 582
Db 466-----VGLASHRELVCYNK----- 480
Oy 583 CELENEKKGRIKLTIEGOLKDLQAKLALSTIPAEKFEKMSKLSNLENEKAKKLDIVEREY 642
Db 481-----TEISNSDLS-----OKLEKTESKY 501
Oy 643 ENSLNETRPLKRELEENLAKLAQHWKPEHEBOLA SRLQONSGELQKRTITELTSKNQLOK 702
Db 502 EELAKVEYLSYQAKM-KLGL-VSPSMDNTHFE-----LKYTE 538
Oy 703 EIEKVCJQDLKLLTOOVNNLTTEKKNPVLKYSSEKMSHVDLYVDLAKKLSDTVHTKYTEK 762
Db 539-----EEIVVLKQDLQNN-LEBSSRNK-----EKV 562
Oy 763 LENEKLLMENASLSKNVSHLETYFIPERPEREKMAKLSNITTELKQLSLKKKQGEDE 822
Db 563 RELERLEVE-----REKGVYAP----- 580
Oy 823 KITSLSMSNNDLKTKSHQGVYKTHEEITLSTLDTRELYOVKXKDCINQEPFK 882
Db 581-----PVEEYEMKSCYGVLENNKKE----- 602
Oy 883 IKDNEBILKRNLENTQNOVKAAYISLRHEEMKSGLSKMKKVOASNAELIAYKKSORE 942
Db 603----- 1002
Oy 943 IYVLEBELAOKRELDITQECIKKAYIISLEDEBKRAKTEKRLKQOILSPOOVXKMS 1002
Db 617 IMKLKO-----TLKSOMTQOE----- 632
Oy 1003 EEEAKKCKOENDKLKEITLQDLKDNKHNLENSYETERALSKNKEELNQMQLNDLQKY 1062
Db 633 SDEA-----EDNKE-----AMNMIDMLKMSDYSOLY 661
Oy 1063 TEAKREKKEYLEENAKOTSEILAOITLOLOVPRDOVYSKLSGCTETLKEKTLQK 1122
Db 663 KEOVALE-----OYRKRSLEDOVTAEYIHKAEHEKLMQILNYSNANK-----EDALSFKMSG- 714
Oy 1123 KQYERBOQOTYOLRQMLENOKNSVFLMLQYAKAEKENGQIKIASKJPEKEESON--- 1179
Db 1180-----VSKYLMETLQDLQDLQAKENSVLITREILQY-----TTLKRLAKMEKESLSLKE 765
Oy 1180-----KTEBYSKQASEIDONTQALKLKEITREYDQLSYKATKASDQLETOISDLNKLAMLR 1235
Db 766 HLAASKVYVAKLEQKLEKKAAM-----TDAMVPSYSEFKLQSSIEBVSYSKLAKSXYK 821
Oy 1236 KYEVOCEVFLIAKKELSLAKODEK---LLHFTSDEKTKOORODSSTLTETLQORIOE 1292
Db 822 EKEKVIHSEYVOIRSEVSOYKREKENIOTLLKSKSEDEYNLQKQAOABEELAMKRRSSS 881
Oy 1293 SAKOIEAKONKTTELILANDYERLQALQNSQLTQSSGSPKRSROSLDLSIOOVNYSLOOQ 1352
Db 882 SKRLEBDEKOKIIBKSEKVTYKIKANLSLQSLST-SISSKRNQSOQDLQDLQOVKQNOLO 940
Oy 1353 LADADROHNOEVIAYIRTHILSLAQHNEDEYQALQLOITOM 1393

```

```

Db 941 LAECKKQNEVISTVYRNHLLAVQOONDEDEVOYKQILITM 981

RESULT 7
ID OPROK7 PRELIMINARY: PRT: 980 AA.
DE 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE NOVEL RETINAL PIGMENT EPITHELIAL CELL PROTEIN.
CN NORPEG.
OS Homo sapiens (human).
OC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxId=9606;
NC K01170:1-9606;
RC TISSUE=RETINAL PIGMENT;
RP RP
SEQUENCE FROM N.A.
RT Kuty R.K., Kuty G., Samuel W., Duncan T., Wiggert B.;
RT "Characterization of a novel gene expressed in human retinal pigment
RT epithelial cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF151135; AF04722.1; .
DR FnuCIG: AF002110; AFK.
DR FnuCIG: AF002110; AFK.
DR SMART: SM00248; ANK_5.
DR SMART: SM00248; ANK_5.
DR PROSITE: PS50088; ANK-REP_REPEAT; 5.
DR PROSITE: PS50297; ANK-REP_REGION; 1.
DR PROSITE: PS50297; ANK-REP_REGION; 1.
SQ SEQUENCE: 980 AA; 110041 MW; 67BA05785D3322 CRC64;

Query Match 17 76; Score 1248.5; DB 4; Length 980.
Best local similarity 26.08; E-value=37;
Matches 370; Conservative 211; Mismatches 333; Indels 507; Gaps 33;

```

```

Oy 17 DMKKYDRLMRAARQDVEKYSILAKKGVNCGKLDVEGSAFHVAVASKQNECTLANILI 76
Db 15 ENMKNDRLQAVNDQAEKVASILGKKGASFTNDSBECKTAFRLMAAKQVCELCAMIT 74
Oy 77 HGVDTITSDRGNALILALNTGHALCLQILQINCPREYVDLQGRFALIDNAACQSS 136
Db 75 HGVDTITSDRGNALILALNTGHALCLQILQINCPREYVDLQGRFALIDNAACQSS 136
Oy 137 IOLDCGASVANDQGRFTPLVLAQWCPRTICQLIDIPRODINSPRONKNTAMISGE 196
Db 133 VOLLEHKSPIMLKDLQGNIPLLAVQNGHSICPELIDROADVNSKSGFTALMACE 194
Oy 197 YCGDAVEVLKNGADVTLILDAQHSYVYARIQDNLITL----- 239
Db 195 IGSNNKVAELIKKQNDLVQSLQVDSQILKSLSEKNGQISLILKSIDQADKTPPKK 254
Oy 240-----KASENSKKQREILAKGPRSLQOORLQSLMDVNTKSNQORHONITOLETEREDL 293
Db 255 QHVOVSKISSEKSGTPYTKAPPPPTSTPDLQVSPSRISYTPSLSG---ESVGFAPPF 312
Oy 294 KERUKTOOPOBILLDKVNGLQLOLNEBVMVADL-----PSEKIKSLAAKERO--- 345
Db 322 KAEIISIREKROILSDSTTG-----ADSLDISBDAODLISLDAKVASLTL 370
Oy 346 HESILPTIEALSKRFKFEPSDHLGSGHFHFKEDM---LLKOGOMINTDSQCTSGNPFVIM 402
Db 371 HNEELQDLQAKSP-KAEADLSFDSYSTRVDLGPRLQKCKGCTSPDSK-SSPSVLIHS 428
Oy 403 OSHSHAPLELALPNOASTSENELIKLELDMARFCDKADRIKONELAIKVAECKAL 462
Db 429 LGMS----- 432
Oy 463 ALCEGVNEDSDQIKQLEDAIKDVQKHWYSESEKVCYKQNHTEHLALKEHLTSDAQTGNHR 522
Db 433-----TTDNVVRIDQLOETLQDLQKLTSSAEKROLO----- 465

```

```

OY 523 LMBELKQDLKQKAYTEGASAEVGRKINQKONENLVEEFKRDCKLMENNRLOKEISM 582
DB 457 -----VELOSRAVELVCLNN----- 411
OY 503 CELEPERKCRKLTEMGQDLDSANLALSTPAEFENKSLLSNLENAKRLIDVERBY 642
DB 472 -----TEISENSDLS-----OQLKEQNSKY 411
OY 643 ESNLEMTRELKRELEMLAKLAOHVKEFEEDLKSLSQKSLGKSLKTELELSKNGTLOK 702
DB 493 EEMAEVLYSVOKOM-----KGLG-VSESNQDYSNHE-----LAVTE----- 529
OY 703 ELKVCYLKRLKTLQOVNLTTEKRVNVLKAVSEEMKSHDVTVDLNRKLSIDVTHKYTEKK 762
DB 530 -----EELNVLKQDLDA-----LEESERNK-----EIV 553
OY 763 LEMKELMNSNLSKNVYRLLEVYFPIPERHEKEDMLKSNITELKQSLSEKKEQEOE 822
DB 554 RELPERKLVF-----REKGVTKRP----- 571
OY 823 KITSLSMENNDRKTSNQVYPAKTHBEIKTALSSTLDTKTNRELVYDKKCEDINOEFVK 882
DB 572 -----PVEEYSEKSSSYSTIENMKMK----- 593
OY 883 IKRHELELKRNLKJTONQVAYETSLHEHEKNSGLKSKKAYQONASLELAKKKSQDE 942
DB 594 -----KALPERKIOGAQBE 607
OY 943 IYTLHEIQAOKRELDITQECIKLKVAPILISECERNKFAKERELEKEOLSOQTOXYNTS 1002
DB 608 IKKLD-----TLKSQTOE-----A 633
OY 1003 EBEAKCKQDNKLKLELTLQDLKQKRVHLENSTETRLSRTEELNRKQLDLQKY 1062
DB 624 SDA-----EIKKE-----KARNIDELKQVSELSQHT 652
OY 1063 TEKKEREKLYEENAKOTSEILAQOTLQKQHPLEQVESYKKSLSSTETLEKLEIKKO 1122
DB 653 KEQAQELE-----DYKRSKSLDVAZYIHAHEHEKMLQKLTNVSAAV-----EOLSEKKSQ- 705
OY 1123 RCTEREGQVQTLQRLQENKSNSSVPLAEHLOVEKAFKEVYGIKASLKEKEESON-- 1179
DB 706 --TSVNLAELOLQVLDQKQNSVSTTEHLDV-----ITTLTAKSEKEEISNLKE 756
OY 1180 ----KTEEYSKQSELOINTQOLAKKLETEBVQDLSYKAVKATLSPQLOSLAMEKLANLR 1235
DB 757 HLASVEVAKLEQOLLEKFAAN-----TDAVPPRSYSEKQLSSLESVVSLAKSEVYK 812
OY 1236 KTEEYCEVYLAKKKELSANDEK-----LHFSIOEIKIDQOEKCKSLTTITIELORIOE 1192
DB 813 EKEKHSVQVQREVSQVREKENIGOTLLKSKEDVBNELLOKFOQAQELAEKKTAVES 872
OY 1293 SAKOIKAKQNTITELLNDYERKQKALNGLSQUTIGGSGPSKRSQSLDLSQOQPSLQOQ 1352
DB 873 SSRLEKDKKIKINEMSKYKIKLAKLSLSQSLST-SSSSKRSQSLQELQOQVQKQDLQNO 931
OY 1353 LADABRQEOVATYRTLLSLAAGIHDEEDVOALLQIOLIM 1393
DB 932 LABCKQHOEIVSYEMHLLLYAVQGDDEVOVKQLQILTM 972
RESULT 8
O9EP71 PRELIMINARY: PRT: 979 AA.
AC O9EP71:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NMRPG-LIKE PROTEIN (ANKYCORBIN).
DE 1700020LIRIK.
OC Eukaryote: Metazoa (mouse)
OC Eukaryote: Metazoa (mouse)
OC Chordata: Vertebrata: Euteleostomi:
OC Eukaryote: Metazoa (mouse)
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

```

```

OX NCBI_TaxID=10090:
RX (1)
RA SEQUENCE FROM N.A.
RA Kutty R.K., Kutty G., Samuel M., Duncan T., Bridges C.C.,
RA El-Sherbeny A., Neglizeni C.N., Smith S.B., Wiggert B.:
RA "Molecular characterization and developmental expression of NMRPG: a
RA novel gene induced by retinoic acid."
RA J. Biol. Chem. 0:0-0(2000).
RX (2)
RA SEQUENCE FROM N.A.
RA STRAIN=C57 BL/6J.
RA Peng Y.-F., Mandel K., Sakitsaka T., Okabe N., Yamamoto Y.,
RA Yokoyama S., Mizoguchi M., Shiozaki H., Monden K., Takai Y.:
RA "Ankyorbin: a novel actin cytoskeleton-associated protein."
RA Genes Cells 0:0-0(2000).
RX (3)
RA EMBL: AF204866; AAC25937.1; -
RA MIM: 602528; AAC25937.1; -
RA MIM: 602528; AAC25937.1; -
RA Interpro: IPR002110; ANK.
RA Pfam: PF00023; ank; 7.
RA PRINTS: PRO1415; ANKRAIN.
RA SMART: SM00248; ANK; 6.
RA PROSITE: PSS0088; ANK_REPEAT; 5.
RA PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
RA SEQUENCE 979 AA; 108852 MW; B2B9C016080237C4 CRC64:
SO
Query Match 17 18: Score 1202; DB 11: Length 979;
Best Local Similarity 24.84; Pred. No. 6.8e-36;
Matches 352; Conservative 207; Mismatches 353; Indels 510; Gaps 28;
OY 17. DNKKYDRLRAHRAERGDEYEVSSILAKGVNPKLDEGSAFHVVASGNLECLMALLI 76
DB 15 ENKNDRLDAENGDNAENYASLKGASNPKHDSGKTFHLLAAGAACCLVAVMT 74
OY 77 HGYDITSDTACNNALILAKTGNALQDLKQLOLTNPTREYVLDQRTALHDMADCCSS 136
DB 75 HGVDYTDQSSGSHALHVAANKGNHBEIKRLQYKSPENINDSKTALHVAAGQDLA 134
OY 137 IQLDCHGASVARDYGRPLVLAITOMKRPITQOILLDGDAGINSRDKONTALMIGE 196
DB 135 VOLDEKSPINIKDIDGNTPLVAVONGHSFACHPLLDNGADVNSBRKRTALMLACE 194
OY 197 YGCKDAEVLKNGADVTLDAIDGDSSTYFARGNDLITLL- 239
DB 195 TGSNVDLIRKQDLSDLSQDNLNHLKSLSEBNAGIQLLSKISQDADLKTTPTPK 254
OY 240 ----KTASENSKREIKKQKQPLQDQNLQSLQDNLNFKNSQENQINQDLEINDL 293
DB 235 QHDVQSTISRSRQTKRKRAAPPISPOLSDVSSPSRITSPISK--ESEVFEMAF 312
OY 294 KERLRKIQEQRILLDKQNGIQLQDLEEVYAADL-----ESEKREKLSLAAEKO-- 345
DB 313 KESLISITQENKDRLSSTAG-----ADSLDSSDAQDQDLVLAQNAISLTL 361
OY 346 HEESIR-TLEAKSREKYPESQHLDSGSPFEREDMLKQCONMTSDQCTSGCPHYNOS 409
DB 362 HKKELQDLQKSPKMKAEADISFQSFQSTQDPLAPGKASDIPSDAKSSPVEH- 414
OY 405 RSLMRPLELADPNQASVSEMLKKELEAMRFPDSAKORLQNLQNELAKHVAECALAL 464
DB 420 ----PAGSTYTDNDVY----- 431
OY 465 ECEYKVEDSDQIKQLELAKQVQRNYSSEKGYQONOTFLAKELHUTSMAQONIRYM 524
DB 432 ----IQLQDLSLQDLQKRLSSAEKQDLQDE-----LQSRDITLCLNMTIS 476
OY 535 E--ELQDLQDKMYVTEGASAEVGRKINQKONENLVEEFKRDCKLMENNRLOKEIS 581
DB 477 ENGSLSQKLEQNSQYEEAKMEVLYSVOKOKL-----GLSQSES----- 516
OY 582 MCELEKREKCRKLTEMGQDLDSANLALSTPAEFENKSLLSNLENAKRLIDVERE 641

```

```

Db 517 -----AACYSHLR----- 525
Oy 642 YERSINIAEPLARELLENKAKIAQVHREPHBQJ.KSLNKQSKOSELQKRNITLTSKNQOTJO 701
Db 526 -----APADEIDITL-----ODJQ 540
Oy 703 KEIKVYCDLKKLLTQOVNLTTEKKKVKLVSEKKSHDYVDDINKLSVTHIKYTER 761
Db 541 KAYE----- 544
Oy 762 KLEMELLIMENASLKKVNSRELVTFIPPERHKKEMALSNITELNKOTSELNKKCEQO 821
Db 545 -----ESNRKNERVELETKL-----AKRQOAEAKR----- 571
Oy 823 EKITYLSMSENNDLKTKSHOVVPKTHETIKTALSSTLTKNRELVADYKCKKEDINOEFV 881
Db 572 -----PACQELASSVCYSTENKMK----- 593
Oy 882 KINDEMLERHLENQONQYAKETLSLGEHEKNSGLKSKKKVQVNSASLAKYKXSE 941
Db 594 -----KATLEKXQAOE 606
Oy 942 EYTLVHEIQAOKRELDITIOCIKLVAPITLSCEKCKFKATKELEQLSQOTQVYNT 1001
Db 607 EIKNKLMD-----TLKSQMPQAP----- 624
Oy 1002 SEEBAKCKOENDKKELITLLONDKCKKVNITPMSEYTEALSRTELSNQLKDIOLK 1061
Db 625 -----DSCQDKKANNRMIDELNKQVSELSQ 651
Oy 1062 YTEKKEKKEVYENAKOTSEILAQOTLLOKQVPLBOVESLKKSLSGTTEFLKEBELKTK 1121
Db 652 YRENAQ-----LEDYRKRSLEDAVYTHRAHEKRLHVSINLSNKS-----EALSLKSKQ 704
Oy 1122 QRCYKREBOQYTOJ.ROMLENQXSSVPLALQYKREKFEYGTITASLEKREPSQ---- 1178
Db 705 -----YSVNLNETLOLKQVDAKHEVNSITHEIQ-----ITTLTAKEHEKEXISALT 754
Oy 1179 -----NNTSEVSKLOSETONTQKALKKLTETREVVDSISKYKATKSULETOISDLEKLEKLN 1234
Db 755 GHILANKEAVAKLEKQALKEERAAV-----SDAMPKSSYERKQASLESEVNAALATKLESY 810
Oy 1235 KRTIEVCEVYLAKKELSAKDEKELHFSI-----SOTIKDOQEKCKSLTTITELORRRO 1291
Db 811 RERRKNSHVAQVRSEVSOAREKNDIOTILKAKQVSEVYALVQKPRQOSELACKRRGCE 870
Oy 1292 ESANQOETAKDNKITELNDYERLKOALNGLSQUTTGSGSPRSOSQIDSLQOVSRLQO 1351
Db 871 TSSLEIEKDEKINEMTREVYKLEKLESLQISLT-TSSSKRSOSQOIDLQOQVKNOLN 929
Oy 1352 GLADADROHQEVTAIYTHLLLSANQCHMEDVOALLQIOTKOR 1393
Db 930 QACCKRHQEVTAIYTHLLLYAVQOCHMEDVOAKYLAQIOTLN 971

```

```

RESULT 9
OQC218 PRELIMINARY; PRT; 227 AA.
ID OQC218
Oy 01-JUN-2001 (TEMBUREL 17, Created)
Db 01-JUN-2001 (TEMBUREL 17, Last sequence update)
Oy 27000059002R1K PROTEIN.
Db 27000059002R1K PROTEIN.
Oy Mus musculus (Mouse).
Db Mus musculus (Mouse).
Oy Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Db Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Oy NCBI_TaxID=10900;
Db NCBI_TaxID=10900;
Oy SEQUENCE FROM N.A.
Db SEQUENCE FROM N.A.
Oy MEDLINE=2108560; PUBMED=1121785;
Db MEDLINE=2108560; PUBMED=1121785;
Oy Kawai J., Shindagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Db Kawai J., Shindagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

```

```

Query Match 14.18, Score 989.5, DB 11: Length 227;
Database Similarity 83.08, Pct Ids 28.2, E-Value 23.1;
Matches 199, Conservative 20, Mismatches 8, Indels 7, Gaps 1;
Oy 1166 IKASLREKREESQKTEVYSKLOSETONTQKALKKLTETREVVDSISKYKATKSULETOISD 1225
Db 1 MKASLREKREESQKTEVYSKLOSETONTQKALKKLTETREVVDSISKYKATKSULETOISD 60
Oy 1236 LNKELANRKYKEVCEVYLAKKELSAKDEKELHFSI-----SOTIKDOQEKCKSLTTITEL 1285
Db 61 LNKELANRKYKIDQCE-----KVASNDKELHLSIDQELROKRECKSLTTITEL 113
Oy 1286 LORRQESAKQTEKDNKITELNDYERLKOALNGLSQUTTGSGSPRSOSQIDSLQOVSRLQO 1345
Db 114 LORRQESAKQTEKDNKITELNDYERLKOALNGLSQUTTGSGSPRSOSQIDSLQOVSRLQO 173
Oy 1346 VRSIQOLADADROHQEVTAIYTHLLLSANQCHMEDVOALLQIOTKORGLVC 1399
Db 174 VRSIQOLADADROHQEVTAIYTHLLLSANQCHMEDVOALLQIOTKORGLVC 227

```

```

RESULT 10
Oy0101 PRELIMINARY; PRT; 1710 AA.
ID Oy0101
Oy 01-JUN-2001 (TEMBUREL 16, Created)
Db 01-JUN-2001 (TEMBUREL 16, Last sequence update)
Oy 01-JUN-2001 (TEMBUREL 17, Last annotation update)
Db BA1458.1 (KIAA1074).
Oy BA1458.1.
Db BA1458.1.
Oy Homo sapiens (Human).
Db Homo sapiens (Human).
Oy Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Db Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Oy NCBI_TaxID=9606;
Db NCBI_TaxID=9606;
Oy SEQUENCE FROM N.A.
Db SEQUENCE FROM N.A.
Oy BLASTED (DEC-2000) TO THE EMBL/Genbank/DBJ databases.
Db BLASTED (DEC-2000) TO THE EMBL/Genbank/DBJ databases.
Oy EMBL: ALI62722; CACI9649.1;
Db EMBL: ALI62722; CACI9649.1;
Oy InterPro: IPR002110; ANK.
Db InterPro: IPR002110; ANK.
Oy Pfam: PF00023; ANK.4.
Db Pfam: PF00023; ANK.4.
Oy SMART: SM00248; ANK.4.
Db SMART: SM00248; ANK.4.
Oy PROSITE: PS50297; ANK-REP_REGION.1.
Db PROSITE: PS50297; ANK-REP_REGION.1.
Oy SEQUENCE 1710 AA; 196410 MW; 01CFB9BADB894872 CRC64;
Db SEQUENCE 1710 AA; 196410 MW; 01CFB9BADB894872 CRC64;

```

```

Query Match 9.38, Score 635.5, DB 4: Length 1710;
Database Similarity 83.08, Pct Ids 28.2, E-Value 23.1;
Matches 199, Conservative 20, Mismatches 8, Indels 7, Gaps 1;
Oy 1166 IKASLREKREESQKTEVYSKLOSETONTQKALKKLTETREVVDSISKYKATKSULETOISD 1225
Db 1 MKASLREKREESQKTEVYSKLOSETONTQKALKKLTETREVVDSISKYKATKSULETOISD 60
Oy 1236 LNKELANRKYKEVCEVYLAKKELSAKDEKELHFSI-----SOTIKDOQEKCKSLTTITEL 1285
Db 61 LNKELANRKYKIDQCE-----KVASNDKELHLSIDQELROKRECKSLTTITEL 113
Oy 1286 LORRQESAKQTEKDNKITELNDYERLKOALNGLSQUTTGSGSPRSOSQIDSLQOVSRLQO 1345
Db 114 LORRQESAKQTEKDNKITELNDYERLKOALNGLSQUTTGSGSPRSOSQIDSLQOVSRLQO 173
Oy 1346 VRSIQOLADADROHQEVTAIYTHLLLSANQCHMEDVOALLQIOTKORGLVC 1399
Db 174 VRSIQOLADADROHQEVTAIYTHLLLSANQCHMEDVOALLQIOTKORGLVC 227

```















Thu, Apr 11 07:43:38 2002

us-09-750-590-2.std.rsp

Page 15

**THIS PAGE BLANK (USPTO)**

OM of: US-09-750-590-2 to: EST:★ out\_format : pfs

Date: Apr 10, 2002 6:36 PM

About: Results were produced by the GenCore software, version 4.5  
Copyright (c) 1993-2000 CompuGen Ltd.

```

Command line parameters:
MODEL=lrme-p2n.model
-0/-cgn.11/0/SPRINT.pool/0/5075030/runat_10042002_145503_10850/amp_query.fasta.1.1480
-MINITERM=1000000000
-SUPPLY=std.rst
-GAPOP=12.000
-GAHEAT=4.000
-DELOP=6.000
-XGAPOP=10.000
-XGAHEAT=0.500
-XGAHEXT=0.050
-XGAPOP=10.000
-XGAHEAT=0.500
-XGAHEXT=0.500
-XGAPOP=10.000
-XGAHEAT=0.500
-DELOP=6.000
-DELETE=7.000
-START=1
-MATRIX=D10sum62
-TRANS=human40.cd1
-LIST=45
-DLOCALIGN=200
-THR_SCORE=100
-THR_MIN=0
-ALIGN=15
-MODE=LOCAL
-OUTFMT=ps
-NORR=off
-HEAPSIZES=500
-MINLEN=0
-MAXLEN=2000000000
-USER=5075030/cgcn1.1.4609
-NCPU=6
-ICPU=1
-LONGLOC
-DEV_TIMEOUT=120
-WARN_TIMEOUT=30
-NO_XLUPY
-NAIT
-THREADS=1

```

Search information block:

Query: US-09-750-590-2

Query length: 1401  
Database: EST.\*

```
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 1392.800000
```

score\_list:

sequence  
qb\_hlc:AK0:[illegible][illegible]

| LOCUS      | 1545 bp                                                                                                       | mrna | HTC | 05-JUL-2001 |
|------------|---------------------------------------------------------------------------------------------------------------|------|-----|-------------|
| AK012454   |                                                                                                               |      |     |             |
| DEFINITION | Mus musculus 11 days embryo cDNA, RIKEN full-length enriched library, clone:2700059D02, full insert sequence. |      |     |             |

VERSION AK012454.1 GI:12849206  
 KEYWORDS CAP trapper.  
 SOURCE Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA,

**ORGANISM:**

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**AUTHORS**  
**TITLE**

JOURNAL  
MEDLINE

**PUBMED  
REFERENCE**

AUTHORS

FILE

MEDLINE  
BIBMED

## REFERENCE

1

11-11-68

MEDLINE  
PUBMED

## REFERENCE

**TITLE**

## REFERENCE

.....

JOURNAL











```

664 1acInHlaVallyserProgluIngluInleuLysSerPrlonglu 680
|||||
101 CTCGACGCTGCAACACGAGGACATGACGAGTAAAGCGATTAA 150
|||||
681 GlnLysSerGlyLysGluLysValGlyThrGluLeuThrSerLysAs 697
|||||
151 CCGAAATGACGAGAACTGGGAGAAAGATGACGATGATTAACATTTG 200
|||||
697 nGlnThrLeuGlnLysGluLysValLysLysLeuLysLeuLys 714
|||||
201 TCAGACGACGAAAGCAATTTGAAAGGATTAATTCGATTATTCCTC 250
|||||
714 eutHrGlnGlnValLysAsnLeuThrThrGlnLysSerLysAsn 728
|||||
251 TCAGACGACGACGACGACGACGACGACGACGACGACGACGACG 300
|||||
729 ProLysValSerGlnLysLeuLysValLysValLysValLysVal 745
|||||
301 CCTTTAAAGTAACTGAGACATGAAACATGATGATGATGATGATG 350
|||||
745 PaPLeuAsnLysLysLysSerAspValThrLysLysValGlyThr 762
|||||
351 TCGATCTTAAAGCAACGCTTAAATGATGACGAAATTAATGACGAA 400
|||||
762 ysrLeuGlnLysGlnLysLeuLysSerGlnAsnLysSerLysSer 778
|||||
401 AGTTGAAATGCAAAATTCCTACGCAAAATGACGCTTAATTAAG 450
|||||
779 ValSerArgLeuGlnLysValPheLeuPheProGluLysGlnLys 795
|||||
451 CTGACGCGCTGACGACGCTGCTGCTGCTGCTGACGACGACGAA 500
|||||
795 uMeGLeuLysLysSerAsnLysThrGlnLysLysGlnLysSerG 812
|||||
501 GATATAGCTCTGAAATTCGATATGCTGAAACGACGCTCTCTG 550
|||||
812 LuLeuAsnLysLysCysGlyLysAspGlnLysLysLysSerLys 828
|||||
551 ACCTTAAAGCAACGCTTAAATGATGACGAAATTAATGACGAA 600
|||||
829 SerGlnAsnAspLeuLysLysThrCysLeuLysGlnLysProL 845
|||||
601 TCTGAAACGCTAACTGAGAGACATGATGATGATGATGATGATG 650
|||||
945 LysThrHlaGlnLysLysThrLysLysSerSerThrLeuAspL 862
|||||
651 TAAACGCTGACGACGCTTAAATGATGACGAAATTAATGACGAA 700
|||||
862 hTrpAsnArgLysLeuValLysLysValLysGlnLysGlnLys 718
|||||
701 CTACACGATTTTACATGCTGAAAGCAATTTTACGATTAATTA 749
|||||
879 GluPheValLysLysLysAspGlnLysGlnLysLysLysThr 895
|||||
750 GAAATTCGTAAGCAATTAATGATGATGATGATGATGATGATG 798
|||||
895 uAsnThrGlnAsnGlnValLysLysGlnLysLysSerLys 908
|||||
799 AAACGCTGACGAAACGATTTAAACGATGATTAATTCCTG 838
|||||

```

```

seq_name: gb_e912.BG176687
seq_documentation_block:
LOCUS      BG176687                1060 bp            mRNA
DEFINITION 60231291f1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4419071 5',
            BG176687
ACCESSION  BG176687.1 GI:12683390
VERSION    BG176687
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE 1 (bases 1 to 1060)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/
INSTRUMENTS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    J. Biol. Chem. 273(19):11921-11927 (1998)
COMMENT    Contact: Robert Rosenberg, Ph.D.
            Email: ccraps@remail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
            DNA Sequencing by: Incyte Genomics, Inc.
            Note: This is a NIH-MGC Library.
            http://imgc.ncl.nih.gov/e/column: 24
            plate: LMU0154
            High quality sequence stop: 739.
FEATURES
    source
        location/Qualifiers
            1..1060
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image="4419071"
            /clone_image="NIH_MGC_95"
            /clone_image="NIH_MGC_95"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: Lymph. Vector: pCMV-Sport6. Site: 1. NotI; Site: 2. SalI. Cloned unidirectionally. oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH-MGC Library."
    ORIGIN
        440 a 170 c 257 g 193 t
        Alignment/scores:
            Quality: 1044.00 Length: 362
            Ratio: 3.503 Gaps: 8
            Percent Similarity: 82.320 Percent Identity: 66.298
        Alignment block:
            US-09-750-590-2 x BG176687
        Align seg 1/1 to: BG176687 From: 1 to: 1060
        1007 LysLysCysLysGlnLysAsnAspLysLysLysGlnLysLeuThrLe 1023
        3 AA
        1023 uGlnLysAspLeuLysAspLysAsnValHisLysGlnLysSerThrL 1040
        53 TCGAAAGATTTGAGATTAAGCAGTCTCATTTGAGAGATCTCATGAA 102
        1040 hrcLysArgLysLysSerArgLysThrGlnLysLysAsnAspGlnLys 1056
        103 TCGAAAGATTTGAGATTAAGCAGTCTCATTTGAGAGATCTCATGAA 1052
        1057 AspLeuLysGlnLysArgLysGlnLysGlnLysGlnLysLys 1073
        153 GACTTGTCCAGAAATTAACGAGAGTAAAGCTGAAAGCTGAAACAC 202
        1073 LcLysLysAsnLysGlnLysSerGlnLysLeuLysGlnLysLeu 1090
        203 AGAAG 252
        1090 eugLysLysGlnLysValProLysGlnLysLysLysLysLysSer 1106
        253 TCGAAAGATTTGAGATTAAGCAGTCTCATTTGAGAGATCTCATG 302
        1107 LeuSerGlyThrHisGlnThrLeuLysGlnLysLysLysLys 1123
        303 CTAAATGCGACATTAAGATTAAGCAGTCTCATTTGAGAGATCTCAT 352
        1123 gcGATTTGGLysGlnLysGlnLysThrValThrGlnLysAspGlnLys 1140
        |||||||

```

```

353 GTCTTAGCGAAGAAAGACGACAGCAGCGACACCAAACTGCATCAATTGTTGG 402
1140 LuAnuGluInValAsnSerSerValProLeuAlaLeuIleuGluInValIys 1156
403 ACAATTCAGAAACAGACCTCTTCGTACCTTCCGACAGCACTTCACGATTTAA 452
1157 GluAlaPheGluLysGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1173
453 GAAAGCATTTGAGAAAAGAACTGGGAATCATTAACCAACCGCTGAGAGAAA 502
1173 sGIuGIuGIuSerGIuAsnLysThrGIuGIuAlaSerLysIleuGlnSer 1190
503 GGAGAGAAAGAGCCCAAAACAAATGAGAAAGTGCATTCCAACCTTCAGTGG 552
1190 LuIleGlnAsnThrLysGlnAlaLeuLysLysGluGluGluThrAlaGluAl 1206
553 AGCTTCAGAAATTCATTAACAAACATTTAAAAATTATAGACGTAAGAGSTTA 602
1207 ValAspLeuSerLysLysLysAlaThrLysSerAspLeuGluThrGlnAl 1223
603 GTTGACCTCTCTAAATATTAAGACCAAAAATGATTTTGGAGACACAGAT 652
1223 eSerAspLeuAsn GIuLysIleuAlaAsn LeuAsnAlaGlyLysTyrGluGI 1239
653 TTTCACTGTTAAACAGAAAGAAATTCGGCAATTTCTGATTAAGAAAGTAAGCA 702
1239 uValGlyGlu GIuValLeuAlaAlaLysLysGIuLeuSerAlaLys 1255
703 ACTATGAGAGAGAAAGTTGGTGATGTGCGAAAAGAGAAATATCTGTGAAGA 752
1256 AspGIuLys GIuLeuLeuIleHisPhe..SerLeuGIuGlnGluIleLysA 1271
753 AATGAGAGAGGAATTCAGTGGAATTTTAAGCAATTCAGACCAACAGATTAAG 802
1271 spGIuGlnGluLysAspLysSerLeuThrThrIleThrGIuGlnGln 1287
803 GATCGGAAGG....GACAGGTGGCATAGCGCCATACACACATTCACA 846
1288 ATGATGATLeuGlnSerLeu.....GCTGAGGAGGAGGAGGAGGAGGAGG 1294
847 AAGAACTCCACCAAGCAAGCAATACACAGAAATGGTGCACCAAGCTTGG 896
1295 .....LysGlnIleGluAlaLysAspAsnLysIleThrGIuLeuLeuA 1309
897 AAGCCAAAGAAATTAAGAAACCGGACCT.....GTA 931
1309 aspGluAlaGlyArgLysAlaLeuAsnSerLysLeuSerGlnLeuThr 1325
932 GACCGTGCCACGACATTAAGAGAGCCCAAGGAGGTGTCCACACACAC 981
1326 TyrGlySerGlySerProSerLysArgIleSerGlnIleIleAspSerLe 1342
982 CACACAGGTGGGACCCCAACAAAGGAAAAACACGTGACGAGACAGTGA 1031
1342 uGlnGlnGlnValArgSerLeuGlnGln 1351
1032 CACACAGAAAGAGCAAAACAAACAGCCG 1059
seq_name: gp_srl1:AA986309

seq. documentation block:
LOCUS AA986309 799 bp mRNA EST 28-MAY-1998
DEFINITION uc79c06.x1 Sugano mouse kidney mRna mus musculus cdna clone
IMAGE1418150.3 similar to TR:028282 Q28282 C3V5 PROTEIN.
sequence:
ACCESSION AF098630.1 GI:3167633
VERSION AA986309.1
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 799)

```

/O/EAD18M1GCS723  
 /Ab\_xref="cazon.10090"  
 /clone="IMAGE:1431850"  
 /clone\_1lb="Sugeno mouse kidney m1a"  
 /sex="female"  
 /dev\_stage="adult"  
 /Ab\_host="DB10H"  
 /note="Organ: Kidney; Vector: pCMVcat-PL3; Site:1; Drail  
 (ACGCTGGG) 2; Drail2 (CAGCATGTG); 1st strand cDNA  
 (ATGGCCGCTTTTTTTTTTTTTTTT); double-stranded cDNA was  
 ligated to a Drail adaptor [TGTGGCCATCATG], digested and  
 cloned into distinct Drail sites of the pME185-PL3  
 vector (5' site CAGCATGTG, 3' site CAGCATGTG). XhoI should  
 be used to excise the cDNA insert. Size selection was  
 performed to exclude fragments <1 kb. The University of Tokyo  
 Institute of Medical Science custom primers for  
 sequencing: 5' end primer CTTCTGCTCAAAATCATGCG and 3' end  
 primer CACGCTCAAGCTGACACACA."

Align seg 1/1 to reverse of: AA986309 from: 1 to: 799

189 r a l a l e u n e t t l e u g l y c s g l u t y r g l y c y s l y s a s p a l a v a l g l u v a l l 206  
|||||:::||||| |||||:::|||||

749 AGCTGTGATGGTCGCTCGGAGTCCGGGTGCAGAGATGCCCTAGACATCC /00  
206 euilelysa nciylala spval ThrleuleuAspa laleugl yHisasp 222

699 TCCTGAAATGCTGCCGACCTGACCTTGCTGTAATGCCCTTGCCACGAC 650  
223 SerSerTyrTylalaarglleclYaspaasnleuasplleuthrleuLe 239

649 AGTTCTACTACGCTAGCATTTGGTGACACCTGACACATCCTGACACTACT 600



```

1190 lulllecllnaanthlysglnalaleuylslysluclugluThaArglUval 1206
|||||
554 AGGTTCAAGATCTAAACACACATTAACCAAGATTAAGAGGCTA 603
1207 ValAspLeuSerLysTyrAlaThrLysSerAspLeuGluThrIleu 1223
|||||
604 GTTCAGCTTCTAAATATTAAGCAACAAAAGGATTTGGAGACACAT 653
1223 eSerAspLeuAsnGluLysLeuAlaAsnLeuAsnArgLysTyrGluLuv 1240
|||||
654 TTCAGCTTAAATGAAGAAATTCGC.AATCTGAATAGAGAGATGACGAG 702
1240 AlLysGlu.GluValLeuHisAlaLysLysGlu 1251
|||||
703 TATCTACGCAAGAGTTTTCATCTCCACAAAACAAAAA 739

seq_name: gb-est2:BC253578

seq_documentation_block:
LOCUS BC253578AF1 893 bp mRNA EST 11-SEP-2001
DEFINITION NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4472692 5',
RNA sequence.
ACCESSION BC253578
VERSION BC253578.1 GI:12763394
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH_MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csparks@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The MGC Consortium (LMNC)
CDNA Library Arrayed by: The MGC Consortium (LMNC)
CDNA Library Arrayed by: The MGC Consortium (LMNC)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNC at:
http://image.llnl.gov
Plate: LM10293 row: p column: 05
High quality sequence stop: 690.
Location/Qualifiers
source
1..893
cdna_construct="Homo sapiens"
db_xref="taxon:9606"
clone_id="IMAGE:4472692"
/clone_id="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/label="DH10B (phage-resistant)"
/notes="Organ: Liver; Vector: pCMV-Sport6; Site 1: NCI;
Site 2: SALT; Cloned using reverse transcript primed.
Full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 305 a 159 c 207 g 212 t

alignment_scores:
Ona lity: 1002.00 Length: 273
Ratio: 4.107 Gaps: 8
Percent Similarity: 89.377 Percent Identity: 81.319

alignment_block:
US-09-750-590-2 x BC253578 ..
Align seq 1/1 to: BC253578 from: 1 to: 883
215 LeuLeuAspAlaLeuGluHisAspSerSerTyrTyrAlaArgLysGluAs 231

```

```

|||||
2 TTCTGATGAGCGCTTGCCATGATGATGTTCTTACATTCGACAAATATGGCA 51
231 PAsnLeuAspIleLeuThrLeuLeuLysThrAlaSerGluAsnSerAsnL 248
52 CAATCTGGACATCTCAACCTTGTTGAAAGATCATCTGCAAAATACCAAC 101
248 ysgLysArgLysLeuTyrLysGlyProSerLeuGluGlnAspLeu 264
102 AAGGAGAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151
265 SerGlnMetLeuAspGluValAsnThrLysSerAsnAlaArgLysGlu 281
152 ACACAGATGACAGATGAAATGCAATGCAATGCAATGCAATGCAATGCA 298
281 nAsnIleGlnAspLeuLysIleLysAsnGluAspLeuLysGluArgLeu 298
202 AATATTTACAGATTTGGATGATTAAGATTAAGATTAAGATTAAGATTA 251
298 rGlySerIleGlnLysGlnArgLysLeuLeuLeuAspLysValAsnGly 314
252 GAAATATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
315 GlnLeuGlnLeuAsnGluLysLysValAlaAspAspLeuSerGln 331
302 CAGTTACAGCTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 351
331 LysGluLysLeuLysSerLeuLeuAlaLysGluLysGlnIleGluG 348
352 GAGGAG 401
348 LysSerLeuArgThrIleGluAlaLeuLysSerArgPheLysTyrPhe 364
402 AAACCTTAAGAGACTATTCAGCTCTGTAAGATTAAGATTAAGATTAAG 451
365 SerArgHisLysSerGlnSerHisPhe.....ArgLysGluAsn 379
452 AGTCATCATTTAGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 501
379 LysLeuLysGlnGlyGlnMetTyrMetThrAspSerGlnCysThrSer 396
502 GCTCTTAAACAGAGTCAGATGTATATGACAGTCAGATGTATATGCTCC 551
396 hGlyMetProValHisMetGlnSerArgSerMetLeuArgProLeu 412
552 CAGGTATACAGCCATTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 601
413 LeuAlaLeuProAsn..GlnAlaSerTyrSerGlnAsn..GlnIleLeu.. 427
602 CTATCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
428 ....LysLysGlnLeuGluAlaMetArgThr..PheCysAsp..SerAla 441
652 CAGCAGACAGAGTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 701
442 ..LysGlnAspArgLysLysGlnAsn..GluLeuAlaHisLysValAla 457
702 AAAACAG 751
457 lGluLysLysAlaLeuAlaLeuGluCysGluArgValLysLysAspSer 473
752 CCAATGCGCAAGAGTTTACGCTTTCATGCTGCGCCAGGCGGAGGCTCC 801
474 AspGluIle 477
802 CATGACAGATTT 813

seq_name: gb-est2:BG120762
seq_documentation_block:
LOCUS BG120762 883 bp mRNA EST 30-JAN-2001
DEFINITION 602348520P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443316 5',
mRNA sequence.

```

ACCESSION BG120762  
 VERSION BG120762.1 GI:12614271  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 663)  
 AUTHORS NIH-NCBI  
 TITLE NIH-NCBI  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cstraub@mail.nih.gov  
 Tissue Procurement: ATCC  
 Library Preparation: Life Technologies, Inc.  
 CDNA Library: Yes  
 DNA Sequencing by: Incyte Genomics, A.G.E. Consortium (LML)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LML at:  
 http://image.llnl.gov  
 Plate: LLM10217 row: h column: 05  
 High quality sequence stop: 669.  
 Features  
 Location/Qualifiers  
 1..883  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="443316"  
 /clone\_id="NIH-MGC-90"  
 /tissue\_type="adeno carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /vector="pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SmaI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size: 1.5 kb  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH-MGC Library.  
 BASE COUNT 345 a 163 c 197 g 178 t  
 ORIGIN  
 alignment\_score:  
 Quality: 995.00 Length: 278  
 Ratio: 3.887  
 Percent Similarity: 92.086 Percent Identity: 83.094  
 alignment\_block:  
 US-09-750-590-2 x BG120762 ..  
 Align seq 1/1 to: BG120762 from: 1 to: 883  
 1090 LauglmglysglnhlyalProleuglmglyalgluSerleuglylyse 1106  
 3 TTGCAGAAACCAACCTTCCTTCATTCGACAGCTGACGCTTCAGAAATTC 52  
 1106 fLeusergylthrlleglulthrlleuglysglnleuglythrllysgln 1123  
 53 TCTTATGTCACACAACTTAAAGCAAGCACTGAACATATTCAGAA 102  
 1123 fGCGYTYTGLySGlmglnthlyalThrlleuglysglnmetleu 1139  
 103 GCGTTCACAAAGAGCAGCAGCACTGACCAATTCGCA TCAATCTG 151  
 1140 GluagngllysaSersevalProleuaglnhlysglnlyal 1156  
 156 GAGATTCAGAAAGCACTCTTACCCCTGCGAGCACTTCAGATTA 201  
 1152 AGTAAATPhgclgyllylletlysaSerleuaglyl 1173  
 202 AGAGCACTTCAGAAAGCAATTCGATCATTAAGACAGCTTGAGAGAA 251  
 1173 yGClugllysergluylthrllyalysrllyserlysglnlyser 1189  
 252 AGGAGAGAAACCAACCAATTCAGCAAGCTTCAGATTCAGCTG 301

1190 Glu1leaglmglnthrllysglnhlyalyllysglnlythrllygln 1206  
 302 GAGTTCAGAAATTCAGCAAGCACTTAAAGCAATTCAGCTGAGAGCT 351  
 1206 lValasleuSerlythrllyalyllyserlyserlysglnlythrl 1223  
 352 AGTGAAGCTTCAGTAAATTCAGCAAGCAATTCAGCTGAGAGAG 401  
 1223 lSerAspLeuagnglylysglnlyalysrllyserlysglnlythrl 1239  
 402 TCTTACCTTAAAGCAAGCAATTCAGCAAGCAATTCAGCAAGCA 451  
 1240 ValCysglglnlyalysrllyserlythrllythrllythrllythrl 1256  
 452 GTATTCAGCAAGCTTCAGTTCAGCAAGCAAGCAATTCAGCAAG 500  
 1256 pGlylysglyleuyllyserlyserlyserlyserlyserlyserly 1273  
 501 TCGAGAGCAATTCAGCTTCAGTTCAGCAAGCAAGCAATTCAGT 550  
 1273 lnglylysglyserlyserlythrllythrllythrllythrlly 1289  
 551 AGGAGCAATTCAGTAAATTCAGCAAGCAATTCAGCAAGCAAG 599  
 1290 lIleaglmgln..SerAllysglnlyleu1alyllyserlyserly 1305  
 600 ATTCAGCAAGCAATTCAGCTTAAAGCAAGCAAGCAATTCAGTAA 649  
 1305 hrclyleuAlasleuAlasleuAlasleuAlasleuAlasleuAlasleu 1320  
 650 CTGACCTTCAGTTCAGTTCAGCAAGCAATTCAGCAAGCAATTCAG 699  
 1321 lLeusergylthrllythrllyser..GlyserProserlythrlly 1336  
 700 CTTCCTCCAACTCAGTAAACAGAGTGGAGCAAGCAAGCAAGT 749  
 1336 rGlnleuyllyserlyserlysglnlyalysrllyserlysglnly 1353  
 750 CCAGCTGATTCAGCTTCAGCAAGCAAGCAAGCAAGTTCAGCAAG 799  
 1353 euAlasleuAlasleuAlasleuAlasleuAlasleuAlasleu 1362  
 800 CTGCAATTCAGTTCAGCAAGCAAGCAAGCAAGCAAGCAAG 828  
 seq\_name: gb\_est1:AV692878  
 seq\_documentation\_block:  
 LOCUS AV692878 705 bp mRNA EST 25-SEP-2000  
 DEFINITION AV692878 GRC Homo sapiens cDNA clone GRC0205 5', mRNA sequence.  
 ACCESSION AV692878  
 VERSION AV692878.1 GI:10294741  
 KEYWORDS EST.  
 ORGANISM human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 705)  
 AUTHORS Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,  
 Xu,S., Li,N., Peng,Y., Liu,F., Ou,J., Song,R., Cheng,Z., Zeng,L.,  
 Gu,Y., Chen,Z., and Han,Z.  
 TITLE Homo sapiens cDNA GK- clones  
 JOURNAL Human Genome Center  
 COMMENT Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzye@hgc.sh.cn  
 This clone is available at GRC in Shanghai.  
 FEATURES  
 Location/Qualifiers  
 1..705



178 AAspApIleAnSerArGAspLySgInAsnArGThAlAldeuKleuEug 194  
194 GCGGAGGTATTATCCAAAGAAACAAAGAAACAAAGTACCCTCAGTAG 51  
2 GCGGAGGTATTATCCAAAGAAACAAAGAAACAAAGTACCCTCAGTAG 51  
194 yCysGlnATrGCTGCGAGATGATGATGATGATGATGATGATGATGAT 211  
52 TTCCCAATATGCTTCGCGAGATGATGATGATGATGATGATGATGATG 101  
211 LaAspAlTrAlLeuLeuAspAlAlEugLyAlAspSerTrTyTAla 227  
102 CTGATATTAACCTTGTGATGATGATGATGATGATGATGATGATGATG 151  
228 ArGAlLeuAspAlLeuAspAlAlEugAlLeuLeuLeuLeuLeuLeu 244  
152 AGATTGCTGACATCTGGACATCTTACACTGTGGTGAAGGTCACATCG 201  
244 uAnSerTrAlLySgLyArGAlLeuArGAlLySgLySgLyPSeArEug 261  
202 AATATACCAAGGAGGAGAACTTTGGAGAGAGAGAGAGAGAGAGAGAG 251  
261 AATATATATATATATATATATATATATATATATATATATATATAT 277  
278 ArGAlAlNAsGInAsnAlLeuAlAspArEugAlAlEugAlAlAspAr 294  
302 AGGAGACCTCAAAATATTCAGATTTGGATTTGAGATTTGAGATTTGA 351  
294 sGAluArGLeuArGAlLySgInGInGAlNAlAlArGAlLeuLeuAsp 311  
352 AGAGAGCTTGGAGAAATATATACCAAGAGAAAGAAATATCTTGGAT 401  
311 AlAsnGAlEugInGInGAlLeuAlAspArEugAlAlEugAlAlAspAr 327  
402 TCATATGATTTACAGTTACAGTTACAGTTACAGTTACAGTTACAGTT 451  
328 LeuAluSerGAlLySgInGInGInGInGInGInGInGInGInGIn 344  
452 CTGGAGATGCGAGAGAAAGAGCTGATCCCTTTGGACACTTAAAGAAA 501  
344 sGAlNAlAsGAlLySgInGInGInGInGInGInGInGInGInGIn 361  
502 GCGACATGAAAGAACTTAAAGAGCATTAAGAGAGAGAGAGAGAGAG 547  
361 ySTyTrPheGAluSerAlPheAlLeuGAlSerGAlLeuAlPheAlGly 377  
548 AATATTATTGAGAGATCATTAAGATGATGATGATGATGATGATGATG 586  
378 AspMetLeuLeuSgInGInGInGInGInGInGInGInGInGInGIn 394  
||||: ||

154 G L A Y G H T h r P r o L e u A l a L a M e t h r G i n n e C S y A a g P r o t h i l l e y 170  
32 G G G G A G A C C A C A C T G T T C T G G C T C T C A G A T A G A G G C C A A C A A T G 81  
170 S G L M e u L e u I l e a s p a g e l y a l a a s p I l e a s p o r t a s p l y g l a 187  
82 T C A A C T C T G A T A G A A G A G A C G A T G T A A T T C A G A C A G A A A A A A A 131

187 snatgthralaleuamelleuenglycylurtyrcllycyslysaapalaval 203  
 132 acgaacagccctcctacgtctagctgcaaaatgctgcagagatcacga 181  
 204 gtuvalaelluallaylaaanglylaaapvalrthleuleuapalaleucl 220  
 182 gaaactttaaattaaaatgactgactatnnaagcttgcctgagccgtgcg 231  
 220 yliaapserseertytyralaargtleglylaapantleuapoleuat 237  
 232 ccagatgattgcttactatgcgaagaattgctgcaaatgcgacattctaa 281  
 237 thleuleuylserthralaasercluaaaseranlyaglyatrgtlualeutr 253  
 282 ccttgctcaaacgtcactgcgaagaaatgcgaacaaagcgaacgaattctgg 331  
 254 lyslyaglyproserleuaglnatrgaaleuaserclmetleuamgcl 270  
 332 aagaaagccgcaactcttggcagacgcaaaattggcagacatgcagatga 381  
 270 uvalaantthlysaerangclnatrgtluahsclnanailedlnaplaug 287  
 382 actnaattgcactgcgaacgcaagcaaaatgcgaacaaatgcgaatgga 431  
 287 luileuagnglylaapleuylscluatrgleuaglyslleglnaglncl 303  
 432 agattgaaaatgagatttgaaagagaggttcagaaaaatttcagacagaa 481  
 304 clnatrgtleuleuaplyvalaanglyleuclnleuaglnleuamngcl 320  
 482 caaaacaaatccttgcgaacaaatgcgaacaaatgcgaacaaatgcgaatga 530  
 320 ugluvalmetvalaapapaleuclusecluyaglyuaglyuys 337  
 531 ggaagcttatgctgctgactgactgcgaacgcaacagaaacgaacgaact 580  
 337 erleuleuallaalygcltlysghlnhslglnslserleuaytrtl 353  
 581 ccccttttgcgaactgaacaaacgaacaaacgaacaaacgaacaaacgaact 630  
 353 eglylaaleuylaserapheylaytrphoclyuaserapheylscl 370  
 631 tgaagctctgcaaaaatgatttactatattt gaaactgactatgcgaact 679  
 370 ercllyserthralaarglysluaapmectleuleuylglnclvclmet 386  
 680 caaaactcactatgcgaacaaacgaacaaacgaacaaacgaacaaacgaact 708  
 387 tyrmethraapserclnclyrthser...Thylmetservalhlsme 402  
 709 tmgctctacacgacagacttggccacttgaagacactccacagatgcag 758  
 402 tclnseratgsermetleuarg 409  
 759 gccatattgaaatncttgcga 780

seq\_name: gb\_cvrl2:BG260897

seq\_documentation\_block:  
 BG260897 974 bp mRNA EST 13-FEB-2001  
 DEFINITION mRNA:250221 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:448041 5',  
 accession BG260897  
 version BG260897.1 GI:12770713  
 keywords  
 source EST.  
 organism human.  
 Homo sapiens  
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
 NIH-MGC http://www.ncbi.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

COMMENT  
 Contact: Robert Streuoberg, Ph.D.  
 Email: c9apba-remail.nih.gov  
 Tissue Procurement: ATCC  
 DNA Library Prepared by: The I.M.A.G.E. Consortium, Inc.  
 Clone sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 plate: L14M10314 row: c column: 02  
 high quality sequence stop: 668.

FEATURES  
 source  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="NIH-MGC.93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /vector="Organ: bladder; Vector: PCMV-SPORE6; Site: 1; Notif:  
 Average Salt: cloned unidirectionally; oligo-dr primed.  
 full-length clones and cDNA library enriched for  
 Note: this is a NIH\_MGC library.

BASE COUNT  
 367 a 165 c 238 g 203 t

ORIGIN

alignment\_scores:  
 Quality: 981.50 Length: 280  
 Ratio: 3.895 Gap: 6  
 Percent similarity: 90.000 Percent Identity: 78.925

alignment\_block:  
 US-09-750-590-2 x BG260897 ..

Align seg 1/1 to: BG260897 from: 1 to: 974

1100 valgluserleuvalyserserleuaglythrlleuclthleuleuylgcl 1116  
 2 gttgagcgtctgcaaaaatccttaattgcgaacaaatgcgaacaaatgcgaacga 51  
 1116 ucluleuylserthlysglnatrgcyltlysluylglnclnlnhvalt 1133  
 52 adaaactgcgaacgactgcaaaagctgctgcgaacaaacgacgacga 101  
 1133 hrcglnleuarglnmetleuagclnlylaaangseraylprdeu 1149  
 102 ccaaaactgcacatttcttgagaaatcgaacaaacaaactcctgaccctg 151  
 1150 alacthlsleuclnvallysgclualahpgecllyugslvalaglylecl 1166  
 152 gcaaaacttgcgaacgactgcaaaagctgctgcgaacaaacgacgacga 201  
 1166 elysalaserleuarglyuglyuglyuglyuglyuglyuglyuglyugly 1183  
 202 aaaaagccgcttgagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaag 251  
 1183 luvalaserleuclnserclulleglnasrthlysgclnalaaleuyl 1199  
 252 aagatgcgaacgactgcaaaagctgctgcgaacaaacgacgacga 301  
 1200 lysleucluthraarglyuvalaapleuaserleuylrthlylaapleuyl 1216  
 302 aaatttgagacagacagagcgtgactctctgttaattatnaacaaacaa 351  
 1216 sserapleucluthrcglnliserapleuamnglylyslleuallaanl 1233  
 352 acttcttgcgaacgacgactgcaaaagctgctgcgaacaaacgacgac 401  
 1233 euasnaarglytyrgclnlyllysgclnlyllyllyllyllyllyllylly 1249  
 402 tgaatgcaaaagatgagacgactgctgcaaaagcttgcgacaaagaa 451

[illegible]





**THIS PAGE BLANK (USPTO)**